Development of CRI-SPA, a mating-based, CRISPR-Cas9 assisted method for high-throughput yeast strain construction, and its applications in yeast cell factory research

Olsson, Helén Emelie

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Development of CRI-SPA, a mating-based, CRISPR-Cas9 assisted method for high-throughput yeast strain construction, and its applications in yeast cell factory research

---

PhD thesis

Helén Olsson

Novo Nordisk Foundation Centre for Biosustainability

Technical University of Denmark

March 2020

Development of CRI-SPA, a mating-based, CRISPR-Cas9 assisted method for high-throughput yeast strain construction, and its applications in yeast cell factory research

Helén Olsson, March 2020

Main supervisor:

Professor Uffe H. Mortensen

Co-supervisor:

Senior scientist Michael Krogh Jensen

Address:

Technical University of Denmark

Novo Nordisk Foundation Centre for Biosustainability

Kemitorvet, building 220

2800 Kongens Lyngby

Denmark

Preface

This thesis serves as partial fulfillment of the requirements of obtaining a PhD degree at the Technical University of Denmark, PhD school of Novo Nordisk Center for Biosustainability. The work presented was carried out in the time period between February 2017 and February 2020 and was supervised by Professor Uffe H. Mortensen at DTU Department of Bioengineering and Biomedicine, which is where the work was predominantly conducted. The work also included an academic exchange with Technical University of Delft between March and April of 2019 and an industrial secondment with River Stone in Copenhagen. The research was funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 722287: PAcMEN (Predictive Accelerated Metabolic Engineering Network).

Helén Olsson, March 2020

Contents

Acknowledgements

This work would not have been possible without the support from a great number of people to whom I would like to extend my gratitude to here.

First, thanks to my supervisor Uffe Mortensen for inviting me to join the Eukaryotic Molecular Cell Biology group as a PhD student. Your enthusiasm for this research and your good sense of humour helped me through many times when failing experiments had its toll on motivation. I have learnt a lot from you and I really appreciate all the time you have devoted to scientific discussions with me during the past three years.

Thanks also to Michael Lisby for hosting me at Copenhagen University whenever I needed to run high-throughput robotic experiments and for teaching me how to deal with thousands of yeast strains at the same time. I would also like to thank Esben Halkjær Hansen and Laura Tatjer Recordá for hosting me during my industrial secondment at River Stone, and Jean-Marc Daran for hosting me during my external research stay at TU Delft – these are experiences that have taught me a lot and that I'm very happy to have had during my PhD. Thanks also to Guldiere Oliveira Brandão – it was great fun to have you as a student and I truly appreciate all your hard work and contribution to this project.

I would also like to express my gratitude to the PAcMEN network. It has been really fun sharing these years together and I would like to thank everyone involved for fruitful scientific discussions and learning experiences but also nice hangouts and traveling memories. Special thanks to Thomas and Anna who made me feel very welcome and at home during my time in Delft.

This PhD would not have been possible without PCR reagents, sterile pipette tips and a whole lot of agar plates. I would like to thank Amalie and Trine for the great support given to me and the rest of the lab people in 223 and for their positive attitude and helpfulness.

I could not have made it through this PhD without sharing my days with some great people that lifted my spirit also in times of stress and struggle. To my lab buddies Tomas, Javier, Wasti, Amalie and Trine: thanks for all your help, for sharing ups and downs and for your company and fun conversations. Thanks also to the nice people in building 223 and a special thanks to my office mates Sidsel and Clara, who I had a lot of fun with.

A PhD is a bumpy ride and moving to a new country without knowing anyone can be hard. I therefore feel lucky to also have met and gotten to know a great bunch of people outside the lab and that I am now able to call my friends. First, I would like to thank my friend and former flatmate Milica, who made me feel at home in Copenhagen and that kept assuring me that "things would be OK" and who was there for me when I was confident they would not. Thanks also to Vasil for being a close PAcMEN ally from the very start and for sharing many beers, laughs and complaints with me. Thanks to Silvia, Sophia, Jenny L and Christine for great talks and for sharing a lot of fun moments with me. Thanks also to Jenny AB – it has been really great having you as a colleague both at Chalmers and DTU, and as roommate in both Copenhagen and Delft. Thanks for all your support and for always offering a listening ear.

Thanks also to the people back home in Gothenburg that continued being a support for me also after moving to Denmark. Annie, Sebastian and Johanna – you have been with me in so many happy but also turbulent times and I am thankful to

have you in my life. Thanks also to my dear friends from Chalmers who helped me not only through university but also these PhD studies: Sofia, Johannes, Anna and Hanna. Thanks to my IndBio/SysBio crew Cathleen, Marlous, Eric and Ben for offering weekend hangouts combined with various discussion on PhD life. Ben, it was also great being your flatmate for the last few months. Your comforting words helped me in times of darkness!

I would also like to extend a special thanks to a few people at the Industrial Biotechnology group at Chalmers who helped me on my academic journey towards this PhD. Thanks to Lisbeth Olsson, Maurizio Bettiga, Cecilia Geijer and Lina Lindahl for offering your support and advice and for being great supervisors, colleagues and scientists that inspired and encouraged me in my decision to do a PhD.

Finally, I would like to thank my family and my boyfriend. Without your support this PhD would not have been possible. Thanks to my mum for all the love and sacrifices, for always looking after me and for your never-ending will to help. Thanks to my aunt Christina and uncle Ingemar for your kindness, for being a second home to me and for being the ones sparking my curiosity and desire to learn in early childhood years. Thanks also to my boyfriend Josh. Things have been a bit upside-down these three years, as we spent most of our time away from each other, but the thought of seeing you during the weekends always kept me motivated. Thanks for being my main support, for always offering advice and for sharing both the good and the bad times with me.

Summary

The yeast *Saccharomyces cerevisiae* is a workhorse of industrial biotechnology and the use of yeast cell factories for the production of industrially important compounds from renewable biomass offers sustainable alternatives to climate-damaging and environmentally polluting oil-based production. Nevertheless, although technological advances in biological engineering is continuously increasing the number of products which can be produced by genetically engineered microbes, the majority of yeast cell factories are not yet commercially competitive with less costly petrochemical production routes. Improving the performance of cell factories in terms of titers, rates and yields is therefore a main goal of metabolic engineering and which is typically pursued by targeted modifications to the native metabolism of the producing organism. However, due to the inherent complexity of biological systems, less rational approaches employing genome-wide screens of large libraries to find genetic variants yielding improved production may be an attractive alternative requiring no a priori knowledge of genetics and metabolism, and enabling the identification of non-intuitive engineering targets. In regards to this, genome-wide mutant strain libraries such as the yeast deletion collection offers exiting possibilities to not only serve as tools in fundamental research but also as screening platforms for metabolic engineering studies. Their implementation does, however, require high-throughput methods for parallel introduction of exogenous genetic material to a large number of individual strains. The research conducted within the scope of this dissertation aimed at developing and establishing such a method, ultimately making strain library screening a more accessible tool in yeast fundamental research and yeast cell factory optimisation.

This thesis first describes the conceptualisation and development of a HT mating-based method for transfer of DNA from a Universal Donor Strain into the mutant strains of the yeast deletion collection. The method, termed CRI-SPA, makes use of CRISPR-Cas9 for integrating the transferred DNA in the genome of the recipient strains, and of Selective Ploidy Ablation (SPA) for haploidising mated cells without the need for meiosis and complex sporulation procedures. The study further presents results of a proof-of-principle experiment in which CRI-SPA was used to transfer a genetic deletion to a subset of the deletion collection and which demonstrated the usefulness of the method in fundamental research uncovering genetic interactions and cellular network structures.

Next, the use of CRI-SPA in cell factory optimisation was established as the HT capacity of the method enabled the yeast deletion collection to serve as a screening platform for metabolic engineering designs. By use of CRI-SPA, genes for synthesising the aromatic, yellow pigment betaxanthin – used here as a proxy molecule for the medicinally important benzylisoquinoline alkaloids - were transferred to the deletion library and mutants with improved production were identified and verified by reverse engineering. Moreover, the study took advantage of the systematic evaluation of phenotypes that the CRI-SPA method enables and an overrepresentation analysis of Gene Ontology terms amongst the top producing mutants yielded observations that could be of interest for further investigation of the translational regulation of the aromatic amino acid pathway.

Finally, the work presented in this thesis also illustrates the possibility of using CRI-SPA for tapping into the great genetic diversity found between different strains of *S. cerevisiae* in order to identify superior hosts for the bio-based production of different compounds. However, strains used in the fermentation industry and strains that have been isolated from the

wild are non-mating and prototrophic and are not compatible with the method. In the presented study, a CRISPR-Cas9 based strategy was used for restoring mating-competence and for ensuring CRI-SPA compatibility in a set of diploid strains of both industrial and environmental origin. The study confirmed the ability of the constructed strains to mate and further demonstrated their use together with the CRI-SPA method. This preliminary study proved the feasibility of the presented approach and encouraged generation of larger strain libraries encompassing greater genetic diversity and which could be a useful tool in the search for strains with improved production capacities.

In summary, the work presented here describes the development of the novel method CRI-SPA for high throughput strain construction and modification of large strain libraries. The dissertation further demonstrates the applicability of CRI-SPA in fundamental science and cell factory optimisation as the method is used in genetic interaction screens, in the search for metabolic engineering targets amongst the yeast deletion library, and in the introduction of genetic material to genetically diverse, diploid strain backgrounds. These examples serve to establish the method and its usefulness and sets the stage for wider application of CRI-SPA in the continued research and optimisation of yeast cell factories.

Dansk resumé

Bagegær *Saccharomyces cerevisiae* er en ofte anvendt arbejdshest I den industrielle bioteknologi og brugen af gærbaserede cellefabrikker til produktion af industrielt vigtige stoffer fra biomasse tilbyder et bæredygtigt alternativ til forurenende oliebaseret produktion, der ødelægger miljøet. På trods af store teknologiske landevindinger inden for biologisk genteknologi har resulteret i et stadigt stigende antal produkter, som kan produceres i genetisk manipulerede mikrober, er det yderst få, der er kommercielt levedygtige i forhold til de tilsvarende petrokemiske processer. Hovedmål inden for "metabolic engineering" er derfor at forøge cellefabrikkernes produktions effektivitet, hastighed og udbytte via introduktion af specifikke genmodifikationer i produktionsværten. Det er dog svært at forudsige, hvilke ændringer, der er gavnlige på grund af biologiske systemers store kompleksitet. Det kan derfor være nyttigt at anvende mindre rationelt baserede metoder, der tager udgangspunkt i screening af biblioteker, der systematisk dækker stammer med mange forskellige genetiske ændringer. Her kan man uden *a priori* viden lede efter genændringer, der kan gavne den proces man er interesseret i, og derved identificere gener, der ikke intuitivt ikke ville kunne findes på basis af eksisterende viden, og bruge dem i nye metabolic engineering strategier. Mutant biblioteker, der dækker deletioner af alle gener i gær tilbyder i denne sammenhæng interessante muligheder, ikke bare som screening værktøj til at finde gener af basal interesse, men også som til at finde gener, der kan anvendes til metabolic engineering. Effektiv udnyttelse af sådanne biblioteker kræver dog "high-throughput" metoder, hvor nyt genetisk materiale kan introduceres ind i de mange enkelte stammer i biblioteket. Formålet med det arbejde, der præsenteres i denne ph.d. rapport, var at udvikle og etablere en metode, der kan gøre det muligt systematisk at introducere et genetisk træk eller en genetisk syntesevej ind de enkelte stammer i store biblioteker screene sådanne biblioteker og dernæst screene kombinations mutanterne for interessante effekter af interesse for basal forskning eller til gavn for cellefabrik optimering.

Denne ph.d rapport beskriver først udviklingen af en high-throughput metode, der muliggøre overførsel af DNA fra en Universel Donor Stamme til de enkelte stammer i et gen-deletions-bibliotek. Metoden kræver ikke transformation, men er udelukkede baseret på, at man krydse gærstammer sammen og lave diploide stammer. Metoden, som er blevet døbt CRI-SPA, anvender CRISPR-Cas9 til at integrere donor DNA ind i genomet på de enkelte bibliotheksstammer, og "Selective Ploidity Ablation" til efterfølgende at haploidisere de diploide stammer uden, at der foregår meiose og sporedannelse. Studierne, der præsenteres her viser "proof-of-principle" eksperimenter, hvor CRI-SPA blev brugt til at introducere en genetisk deletion i en del af deletionsbibliotket. Resultaterne demonstrerer tydeligt anvendeligheden af CRI-SPA metoden til at identificere genetiske interaktioner i cellulære netværksstrukturer.

CRI-SPA metodens anvendelighed til brug i metabolic engineering er også blevet etableret, idet metodens high throughput kapacitet blev brugt til at screen hele deletionsbiblioteket for mutanter, der gavner produktion af farvestoffet betaxanthine. Dette aromatiske stof fungerer som en proxy til detektion af mutanter, der kan fremme produktion af alkaloidet benzylisoquinoline, der har stor medicinsk anvendelse. Adskillige interessante gener blev identificeret i dette scree; og de bedste blev valideret via reverse genetic engineering, Endvidere, studiet drog fordel af en systematisk evaluering af de CRI-SPA genererede mutanter og en overrepræsentations analyse af Gene Ontology termer blev udført. Dette afslørede også observationer, der kan bruges fremadrettet til at studere translationel regulering af syntesevejen til produktion af aromatiske aminosyrer.

Endelig bliver det beskrevet, hvordan CRI-SPA kan bruges til at udnitte den naturlige gærstamme diversitet til metabolic engineering med henblik på at identificere den bedste mulige gærstamme baggrund til bio-baseret produktion af et givent produkt. En særlig udfordring udgøres af industrielle stammer og vildtype stammer, der er prototrophe og ofte ikke kan krydses. Sådanne stammer er ikke umiddelbart kompatible med CRI-SPA. I de tilfælde bliver man nød til at manipulere stammerne genetisk, før de kan anvendes til CRI-SPA eksperimenter. I denne rapport udvikler vi et sæt CRI-SPA kompatible diploide stammer, hvor stammerne enten er hentet fra naturen eller fra industrien. For at opnå dette har vi gjort dem krydsningskompetente og auxotrophe ved at indføre nogle få genetiske ændringer via CRISPR-Cas9 teknologi. Endvidere blev det vist, at de efter disse ændringer kan bruges til CRI-SPA eksperimenter. Dette lovende præliminære studie har vist mulighederne for at bruge CRI-SPA i disse mere komplekse systemer og opmuntret til at lave større samlinger af CRI-SPA kompatible stammer. Sådanne samlinger skulle være et glimrende startpunkt for at udnytte den store genetiske variation, der findes blandt gærstammer til udvikling af gode cellefabrikker med forbedret produktions kapacitet.

Tilsammen beskriver studiet udviklingen og karakterisering af en ny metode, CRI-SPA, der kan anvendes til high-throughput stamme konstruktion med udgangspunkt i store eksisterende gærstammebiblioteker. Arbejdet demonstrerer endvidere anvendeligheden af CRI-SPA metoden både til bidrage til basal videnskab og til at optimere cellefabrikker idet metoden kan bruges til at screene for nye gener der kan indgå i metabolic engineering strategier og til at screene genetisk forskellige diploide gærbaggrunde for optimale produktionsværter. De viste eksempler er blevet brugt til at etablere metoden og demonstrere dens anvendelighed og sætter dermed scenen fro at CRI-SPA den kan tages i brug i en lang række applikationer i den fortsatte forskning efter ny viden og i konstruktionen af forbedrede cellefabrikker

List of abbreviations

5-FOA	5-Fluoroorotic acid
BIAs	Benzylisoquinoline alkaloids
BIR	Break induced replication
CRI-SPA	CRISPR-Cas9-Selective Ploidy Ablation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DSB	Double strand break
DSBR	Double strand break repair
FACS	Fluorescent activated cell sorting
G protein	Guanine nucleotide-binding protein
GHG	Greenhouse gas
GO	Gene ontology
GPCR	G protein coupled receptor
gRNA	Guide RNA
HR	Homologous recombination
HT	High-throughput
HTS	High-throughput screening
L-DOPA	L-3,4-dihydroxyphenylalanine (L-DOPA
LOF	Loss of function
MAPK	Mitogen-activated protein kinase
ORF	Open reading frame
L-Phe	L-Phenylalanine
SDSA	Synthesis dependent strand annealing
SGA	Synthetic genetic array
SPA	Selective ploidy ablation
SSA	Single strand annealing
THF	Tetrahydrofuran
L-Trp	L-Tryptophan
L-Tyr	L-Tyrosine
UDS	Universal Donor Strain

Introduction

In order to prevent devastating effects on climate and the environment, our society needs to urgently cease its use of fossil resources and find alternative routes for producing the energy, fuels, chemicals and materials of which it is dependent on. In industrial biotechnology, microorganisms are used to convert a renewable substrate into a product of interest. By the use of genetic engineering and expression of heterologous genes, these microorganisms that are typically referred to as cell factories, can be made to produce compounds which are non-native to their metabolism and that have great industrial value. The yeast *Saccharomyces cerevisiae* is an important production host in industrial biotechnology, where it for a long time has been used in the production of bioethanol and therapeutic proteins, and much research is now also devoted to the development yeast cell factories for a wide list of other chemicals.

In order to ensure a sustainable and economically competitive manufacturing process, the native metabolism of yeast cell factories are altered through metabolic engineering to optimise production of the target compound in terms of titers, rates, and yields. This field of research has been greatly advanced by recent technological developments. Tools in systems biology has enabled complex biological systems to be studied and simulated using mathematical models to predict the phenotypic effects of different genetic designs, and gene editing technologies such as Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 has greatly increased the speed and precision of which such genetic designs can be introduced in the yeast cell factory hosts. These and other combined efforts in advancing the field of biobased-based production, has enabled microbial cell factories to produce a wide range of commodity and speciality chemicals, of which a growing number can now be produced at commercial scale. Nonetheless, the great majority of yeast and other microbial production systems are not providing titers, rates or yields that are economically competitive with current oil-based manufacturing processes. Moreover, the establishment of a new and efficient cell factory is a lengthy process, requiring many years of research that does not align with the urgent need for sustainable alternatives to replace climate-damaging oil-based production.

Because of the complex genetic networks and multi-level regulatory mechanisms of biological systems, prediction of phenotypes from genetic designs remains a challenge and metabolic engineering strategies do not always yield the expected improvements to the phenotype in question. For these reasons, screening large genetic libraries to identify engineering targets for improved production of a compound of interest, may be a powerful alternative. Such non-targeted approaches requires no *a priori* knowledge about cellular network structures and enables identification of targets which effect on the studied phenotype may seem unintuitive. In yeast cell factory research, such screening approaches are possible for example by the availability of systematically constructed mutant strains libraries. The most well-used of these libraries, the yeast deletion collection, contains a set of ca. 4800 haploid yeast strains amongst which each non-essential gene has been deleted. The use of these libraries thereby allows genome-wide, systematic screens for elucidation of genotype-phenotype relationships, so that for example the individual effect of deleting each non-essential gene on the production of a target compound can be studied. However, the use of such multimember strain libraries necessitates the use of high-throughput methods for transformation. The natural ability of yeast of opposite mating types to mate to each

other offers a simple system for transferring genetic features harboured within one strain, into a library of many strains by parallel mating. High-throughput mating-based methods for modification of the yeast deletion collections exist but do, however, currently require the use of complex meiosis and sporulation protocols, which limits their applicability and usefulness in yeast cell factory optimisation.

Project aim and thesis outline

To facilitate the use of systematic strain libraries in the search for improved yeast cell factory designs, this work aimed to develop a method that combined the simplicity of yeast mating for manipulation of a large number of strains, with the versatility and efficiency of CRISPR-Cas9 technology, while simultaneously avoiding the need for meiosis and sporulation. To this end, the novel, high-throughput strain construction method CRI-SPA was developed. This work further aimed at establishing the usefulness of CRI-SPA in cell factory optimisation by demonstrating application of the method in both fundamental and applied research questions.

Chapter 1 of this thesis offers theoretical background relevant to the field of study. **Chapter 2** describes the conceptualisation and development of the CRI-SPA method while demonstrating its functionality and applicability in genetic interaction screens using the yeast deletion library. In **Chapter 3**, the possibility of exploiting the yeast deletion collection for metabolic engineering via use of CRI-SPA is demonstrated. The method is used to transfer a biosynthetic pathway for betaxanthin pigment production to the ca. 4800 mutant strains of the collection, which are then screened for their ability to produce the compound. In **Chapter 4**, the versatility of CRI-SPA is demonstrated as we show expansion of the method to be used with libraries of strains of various genetic backgrounds and which allows exploitation of genetic diversity in the optimisation of yeast cell factories. In **Chapter 5**, a discussion on the presented work is given, evaluating the ability of CRI-SPA to meet the aim of this thesis and outlining future directions for the method. Concluding remarks on the technological development in metabolic engineering and microbial cell factory research are also presented.

CHAPTER 1: **Theoretical background**

Our petrol-based society and changing climate

The industrial era began around the start of the 1800s and was then accompanied by the onset of a large-scale consumption of carbon-dense fossil materials such as oil and gas for the production of fuels, energy and goods. The use of these fossil resources also enabled atmospheric nitrogen to be used in the production of fertilisers at an industrial scale, thereby contributing to the elimination of constraints previously put on food production (1). With improvements to human welfare and food supply, the world's population quickly grew, particularly after the Second World War. The one billion people inhabiting the planet at the start of the 19th century has expanded to stand today at 7.7 billion and is projected to grow to 9.7 billion by the year 2050 (2). In parallel with population growth, the number of products made from petroleum increased rapidly and, with an ever growing economy, so did the consumption of these goods. During the year of 2019, global consumption of crude oil reached 100 million barrels of oil per day (3), reflecting our society's heavy dependency on fossil resources for its production of energy, transportation and heating fuels, bulk and specialty chemicals, pharmaceuticals, and synthetic materials including plastics and rubber**.** The fast, fossil-driven industrialisation and expansion of economies and human welfare (at least in Western countries) has, however, not come without a cost.

Fossil materials represent a store of carbon from organic matter exposed to massive geological forces during tens of millions of years. Combustion of these fossil fuels and its derivatives releases this previously fixed carbon to the environment as CO_2 (among other emissions). Therefore, with the large-scale use of these fossil resources, mankind has rapidly (relatively speaking) increased the atmospheric concentration of CO_2 so that it is now at its highest since the onset of the Holocene (1) – the current geological epoch dating back 11 700 years. As the incoming solar radiation is absorbed by the surface of the Earth, thermal radiation from the Earth itself is generated at wavelengths that are efficiently absorbed by CO_2 and other greenhouse gases (GHG) in the Earth's atmosphere, thereby trapping most of the radiating heat. The warming effects of increasing atmospheric concentrations of GHG have been known for a long time (4) and in the past decades it has been made clear that there are severe risks to society and the planet associated with a warmer climate, including rising sea levels due to melting ice caps, severe draughts, an increase in extreme weather events, collapsing ecosystems, and resulting geo-political conflicts (5).

Because of the threats associated with a changing climate, the great majority of the world's countries have now signed and ratified the United Nations' Paris Climate Agreement negotiated in 2015. The nations are thereby pledging to decrease GHG emissions in order to keep global warming below 2 °C compared to pre-industrial temperatures, and to pursue significant efforts to limit warming to 1.5 °C. Warming of 2 °C or more is expected to result in significant negative impacts on ecosystem services and human health in future generations. In order to reach the goal of the Paris Agreement, urgent and resourceful action is needed. In the latest 2018 special report released by the Intergovernmental Panel on Climate Change (IPCC), it was estimated that global emissions need to be cut by 45 % from 2010 levels by the year 2030, in order to have a reasonable chance to stay below 1.5 °C of warming (6).

The biobased economy and industrial biotechnology

In order to prevent the devastating effects of an increasingly warmer climate, our societies need to urgently transition to an economy that instead of using fossil materials, uses renewable resources such as crops, forests and animals and other living organisms, to produce not only food but also materials, chemicals, fuels and energy. Such a system is referred to as a biobased economy or bioeconomy. In 2016, the combined bioeconomy within the EU was estimated to have a turnover of €2.3 trillion (numbers including agriculture, forestry, and the food and beverage industry) (7). A smaller and more specific sector within the bioeconomy is industrial biotechnology. This is the discipline of using microorganisms and/or enzymes derived from these organisms, to produce chemicals, fuels and pharmaceuticals. In some cases, enzymes are also the product itself. Industrial biotechnology has been identified as a Key Enabling Technology by the European Commission for securing innovation, competiveness and growth within the EU (8) . A 2016 report from the European Association for Bioindustries further estimated that the full value chain of the industrial biotechnology sector in Europe employed ca 486 000 people and generated more than €31 billion in turnover in 2013 (9).

Examples of products manufactured by the industrial biotechnology sector include enzymes such as lipases, proteases and amylases used in the food, feed, and chemical industry (10); amino acids for food and feed supplements (11); bioethanol for the biofuel industry (12); bioplastics (13); antibiotics (14); therapeutic proteins (15); and industrial chemicals such as butanol (16), succinic acid (17) and 1,3-propanediol (18). Microorganisms from across the tree of life are being harnessed in these processes, including yeast, filamentous fungi, bacteria, and algae; each with their own benefits and pitfalls. Still, industrial biotechnology is faced with multiple challenges, a significant one being the continuously low price of oil, which makes competition with the petrochemical industry difficult. Additionally, in order to be cost-effective and sustainable (as in non-competitive with food production and forest preservation), biobased production should ideally operate using cheap, renewable feedstocks derived from waste streams such as forestry and agricultural residues. This is a considerable challenge and requires not only efficient infrastructure and logistical networks for the collection, storage and transportation of the material (19), but also considerable technological know-how regarding the processing of such materials. Another challenge is to identify or generate microorganisms with an ability to use an expanded substrate range to utilise carbon sources from lignocellulosic material, municipal waste and even CO_2 (20). Furthermore, to generate yields, concentrations and productivities that allow biobased processes to be competitive with the petrochemical industry, it is crucial that the performance of the producing microorganisms is improved. This is a central goal of metabolic engineering and much research conducted within the field of industrial biotechnology has, and is, being devoted to improving the metrics of microbial production systems (covered further in a later sections).

Saccharomyces cerevisiae – model eukaryote and industrial workhorse

One of the most well-used microorganisms in industrial biotechnology is the yeast *Saccharomyces cerevisiae* - a unicellular organism belonging to the kingdom of fungi and also referred to as budding yeast, baker's yeast, or simply yeast. *S. cerevisiae* is amongst the best-studied and well-characterized eukaryotic organisms in the world and humans and yeast share a long history. Records from ancient China suggest that already some 9000 years ago (21) humans were, although unknowingly at the time, using the aptitude of yeast to convert sugars contained in fruits and cereals into ethanol and other flavoursome compounds appreciated by the human palate. For millennia, yeast has then been used

in the production of alcoholic beverages and in the raising of baked goods. Today, *S.cerevisiae* and other closely related yeasts within the *Saccharomyces* genus are of central importance to the multi-billion dollar baking, brewing and winemaking industries (22).

In the past decades, *S. cerevisiae* has also become a widely popular model organism for studying genetics, physiological processes and disease (23). Its popularity has numerous reasons. Firstly, yeast can easily be grown in the lab using non-expensive substrates and under adequate conditions, the organism displays a relatively fast growth with a doubling time of around 90 minutes. Furthermore, yeast is genetically tractable and a number of well-established molecular biology tools and protocols are available for its facile genetic modification (24). Due to its long use in traditional fermentation processes, *S. cerevisiae* has also been given the GRAS (Generally Regarded as Safe) status by the U.S Food and Drug Administration. In 1996, *S. cerevisiae* further became the first eukaryotic organism to have its genome fully sequenced, revealing a genetic set-up of approximately 6000 genes (25). Importantly for the yeast research community, the sequences and functions of these genes are freely accessible through the curated Saccharomyces Genome Database (SGD). As the human genome sequence later became available, it was found that approximately 60 % of *S. cerevisiae* genes share significant homology with human genes (26) and studies in yeast has been pivotal for understanding the function of a significant number of these. Key to the success of many of these studies was the generation of the yeast gene deletion collection in 1999. This collection, which is covered in greater detail in later sections, is yet another important resource available to yeast researchers. Consisting of ca. 4800 strains derived from a common strain background in which each non-essential gene has been deleted, this collection has been a crucial tool for enabling high-throughput functional genomic studies in yeast (27).

In more recent years, *S. cerevisiae* has, in addition to its role as a model eukaryote, also become an important host for the biotechnological production of a range of industrially relevant chemicals and important medicinal compounds. *S. cerevisiae* is an attractive choice not only because of the great amount of accumulated knowledge about the organism and its genetic tractability, but also because it displays robust growth in industrial settings and a high tolerance to ethanol and acidic conditions (28). In 1987, Novo (nowadays Novo Nordisk) launched their industrial production of human insulin using a recombinant strain of *S. cerevisiae,* replacing the extraction of insulin from porcine pancreas**.** Since then, yeast has been used to produce a number of pharmaceutical proteins, including human serum albumin and vaccines against hepatitis and human papillomavirus (HPV) (15). As the need for non-fossil production processes has become increasingly clear, *S. cerevisiae* has also become a popular host for microbial production of a range of different chemicals and fuels (28). These include native host products such as ethanol (29) for the biofuel industry and important precursor chemicals such as the organic acids succinate (30) and pyruvate (31) but also heterologous compounds produced by expression of heterologous genes and pathways. In this scenario, the yeast is often referred to as a cell factory, which metabolism is “hijacked” and altered by metabolic engineering in order to convert a sugar-based substrate into the product of interest through a cascade of enzyme reactions hosted within the cell.

Metabolic engineering of cell factories – successes and challenges

In industrial biotechnology, yeast and other microbial cell factories convert a renewable substrate into a product of commercial interest. These products may be derived from the native metabolism of the cell factory host or it may be a heterologous compound produced by means of heterologous gene expression. However, production levels in the employed background strains are typically low. In addition to introducing the biosynthetic genes needed for production of the target compound it is therefore also necessary to modify the metabolism of the host to optimise its production in order to enable a sustainable and economically competitive manufacturing process. This field of research is termed metabolic engineering and typically employs a strategy in which cell factory host genes are either deleted, repressed or overexpressed in order to divert metabolic fluxes to the desired product (32). Ultimately, metabolic engineers strive to improve the production metrics of the cell factory in terms of titre (product per volume), rate (product over time) and yield (product per substrate) (TRY, Titre-Rate-Yield). There is typically a trade-off between these metrics and depending on the substrate, the fermentation process, and the final product, it is important to maximise one or the other.

Technological development has provided metabolic engineering with an increasing number of tools to study and to improve the performance of cell factories. Today, metabolic engineering is aided by systems biology, which provides computational techniques for studying complex biological systems. By generation of genome-scale metabolic models (GEMs) that represent mathematical reconstructions of cellular metabolic reactions, *in silico* predictions of behaviours (phenotypes) of the modelled system are enabled. The first GEM for *S. cerevisiae* was created in 2003 (33) and is continuously curated and updated and the most recent version, Yeast8, was released in 2019 (34). Today, genetic design can also be implemented *in vivo* more quickly than ever before, as the introduction of new gene editing technologies such as CRISPR-Cas9 (covered in later sections) has greatly improved the speed and accuracy of genetic engineering. Furthermore, the continued development and advance of synthetic biology is providing metabolic engineering with non-natural DNA constructs, gene circuits and biosensors that enable fine-tuning of cellular metabolism (35).

Metabolic engineering of yeast cell factories has enabled the production of a wide range of biochemicals (reviewed here (28)) with noteworthy examples including medicinal molecules such as the antimalarial drug artemisinin (36) and opioid-derivatives thebaine and hydrocodone (37); biofuels such as biobutanol and biodiesel (38,39), bulk chemicals for the chemical industry such as 1,2-propanediol and lactic acid (40,41); and fine chemicals such as vanillin and resveratrol (42,43) for use in the food and nutraceutical industry. A growing number of these yeast cell factories and other microbial production systems are now also able to operate at a commercial scale (44,45). However, despite successes and technological advance, the great majority of microbial cell factories are still unable to deliver the titer, rates or yields that would allow economic competitiveness with current petrochemical production processes, especially when considering the production of low-value, high-volume bulk chemicals. Furthermore, the generation of new cell factories are hindered by lengthy development processes requiring many years of research, to some degree relying on a trial and error approach in which many genetic designs need to be conceptualised, constructed and tested (“design-build-test” cycles) before an efficient cell factory can be obtained.

Accurate prediction of phenotypes from genotypes remains a central challenge and bottleneck in metabolic engineering (32). Biological systems are inherently complex and our understanding of them are far from complete. Intricate genetic networks and interactions makes it difficult to accurately predict the effect that the modification of individual components have on the system as a whole. Additionally, there are still unknown genes and cellular processes whose effects cannot be predicted nor included in a metabolic genome scale model. Evolution has further optimised biological systems to maintain metabolic homeostasis and complex regulatory mechanisms on multiple levels may efficiently counteract the intended effect of a metabolic engineer design. Lastly, it is also difficult to predict the effect of the metabolic burden that expression of a set of heterologous genes places on the production host by draining the host of biochemical building blocks and ATP, and this over-all burden may impose hard-to-deduce constraints on the synthesis of the intended product (46).

Alternative routes to improved cell factory performance

As outlined above, engineering a biological system to perform in a predictable manner is a considerable challenge, and when targeted rational metabolic engineering approaches fail to deliver the expected improvement, other less targeted screening procedures may be employed at a larger, often genome-wide scale. These screening procedures are powerful as they require no *a priori* knowledge about cellular networks and regulatory mechanism and they enable identification of engineering targets that may be unintuitive from a rational metabolic engineering perspective.

Adaptive laboratory evolution

Adaptive laboratory evolution (ALE) is an alternative strategy to rational metabolic engineering whereby one employs cycles of either natural or induced mutagenesis to the production host, followed by selection for cells with desirable phenotypes. The strategy is a powerful tool and has frequently proven successful for improving various aspects of yeast cell factory performance including for example increased tolerance to elevated temperatures (47), improved tolerance to stress (i.e. robustness) and utilisation of xylose in lignocellulosic fermentations (48), and complete re-wiring of the yeast metabolism abolishing ethanol fermentation to increase flux towards free-fatty acid synthesis (49). When evolved strains are sequenced, shared mutations amongst the top performing strains can be identified and used to reverse engineer the original producing strain in a conventional manner in order to confirm the phenotype. However, when mutations are spread across the genome, deducing which genetic changes are responsible for the improved phenotype may be a challenge. In these cases, evolutionary engineering offers limited understanding of the underlying metabolic networks. An obvious limitation of ALE is that it can only be used to improve phenotypes for which an efficient selective pressure can be applied. This is often not possible if the aim is to increase the production of a compound that is decreasing the competitive fitness of the producing strain. For this reason, much attention is now devoted to the development of different biosensors, designed to detect various compounds of interest and activate reporter genes producing an easily measureable output (typically a fluorescent signal) in a dose-dependent manner, so that evolved cells can be more easily screened (50). Alternatively, biosensors can be designed to, upon detection of a target molecule, drive the expression of a gene needed for survival providing high-producing cells with a competitive advantage.

Library screening

Another strategy employed in cell factory design and optimisation is to screen large libraries of DNA, such as overexpression (51), genome-shuffling (52), random mutation (53) and different cDNA libraries (54) in order to find designs that lead to improved production of a target compound. The libraries are typically in the form of plasmids, which are pooled and transformed into a tester strain. Top producing cells are then identified and isolated and the sequence of the plasmid is obtained in order to reveal the beneficial mutation/component.

As addresses in earlier sections, the yeast genome-wide mutant deletion collection represents an important resource in yeast genetic research. However, although the collection has been used in screens investigating the role of different genes in tolerance to a number of inhibitory compounds found in industrial fermentations (55–57), it has only sparingly been used to screen for differences in the production capacity for different biochemicals. Using the library to identify mutants with improved production of a heterologous compound presents a particular challenge, as it requires all the strains in the deletion library to be transformed with heterologous genes for enabling its production. Conventional yeast transformation is a time-consuming and cumbersome process and might explain the limited number of studies screening the deletion collection for improved cell factory performance. However, alternative methods for introducing genetic modifications to the mutants of the deletion library exist and are described below.

The genome-wide yeast mutant libraries and methods for their utilisation

After completion of the yeast genome-sequencing project in 1996, there was a need to study and uncover the function of the newly discovered open reading frames (ORFs) in the *S. cerevisiae* genome. These studies were greatly enabled by the following creation of the yeast deletion collection - a set of mutant strains in which each non-essential gene was systematically deleted by replacement of the ORF with the selectable *kanMX* cassette (58,59). These mutant strains are derived from a common S288C genetic background and the library exists in haploid versions of both mating types, and as homozygous and heterozygous diploid versions. The collection, which was completed in the early 2000s, started a new era of functional genomics, enabling systematic studies of gene functions and simplified exploration of gene-environment interactions and gene-drug interactions (27). As the mutant strains can be arrayed in and ordered fashion on agar plates, phenotypes can easily be scored and ascribed to specific gene deletions. Additionally, each strain also carries its own unique DNA barcode, allowing the whole collection to be analysed in a pooled format. In this case, pooled strains are exposed to growth conditions causing competition between the different mutants and the abundances of individual barcodes are analysed via microarrays or next generation sequencing (60,61) to reveal mutations deleterious or beneficial to growth under the specified condition.

In later years, the genome-wide deletion collection has been complemented with additional libraries that also enable systematic studies of essential *S. cerevisiae* genes. These include collections of mutant strains expressing hypomorphic alleles (62), temperature sensitive alleles (63), or native genes under the control of the repressible Tet promoter (64). Other types of genome-wide libraries are also available, including for example the plasmid-based Tiling collection for overexpression screens (51) and the GFP tagged (65) and TAP tagged (66) ORF libraries for protein localization and expression studies, respectively. Moreover, *S. cerevisiae* deletion libraries also exist for a prototrophic version of the

original haploid background (67), for the filamentous Sigma 1278b genetic background (68), and for the haploid wine strain AWRI1631 (library construction partly described by (69). Finally, deletion collections are also available for other microorganisms including *Schizosaccharomyces pombe* (70), *Escherichia coli* (71), part of the *Candida albicans* genome (72), and additionally, a large-scale transposon-based disruption library is available for *Yarrowia liplolytica* (73).

Although single gene disruptions are useful for functional analyses, biological phenomena can rarely be ascribed to the function of an individual genetic element and genes and their products are not isolated units but rather part of complex genetic networks. Therefore, an important contribution to the field of functional genomics was the introduction of the Synthetic Genetic Array (SGA) analysis in the early 2000s (74). The use of SGA allowed a second loss of function (LOF) mutation to be introduced to the mutants of the deletion library, thereby enabling genetic interactions to be investigated. These studies are powerful because they can reveal underlying structures of genetic networks and interaction between pathways, ultimately helping resolve the metabolic map of the cell. The principle behind the SGA method is briefly described in the following section. Since the introduction of the method, a handful of other methods for utilising and/or manipulating the yeast genome wide mutant collections have also been developed. These include the Selective Ploidy Ablation (SPA) method , which is thoroughly described in Chapter 2 of this thesis, but also a number of other tools that will not be covered here but which include for example the dSLAM (75) method and the so called Green Monster method (76).

Synthetic Genetic Array (SGA) analysis

To construct double mutants, the SGA method makes use of a haploid *MATα* strain harbouring a query mutation linked to a dominant selectable marker that is transferred to the strains of the deletion library by mating in arrayed manner on agar plates. The procedure is schematically outlined in Figure 1. After crossing, resulting diploids are sporulated and haploid cells of the correct double mutant genotype are generated by a series of transfers onto different selective plates.

First, sporulated cells are cultivated on medium lacking histidine. Since the mutant query strain harbours a *HIS3* gene under control of the *MATa* specific promoter of *MFA1*, whereas deletion mutants are auxotrophic for the amino acid, this step allows selective germination of *MATa* progeny only. Haploid cells are thereafter transferred onto medium supplemented with G418 (selecting for the *kanMX* marker of the deletion mutants) and a second antibiotic to which the selectable marker of the query mutation confers resistance to. In order to handle the great number of strains in the deletion library, the method requires the use of automated instruments for pinning and replicating the arrayed strains onto the different plates.

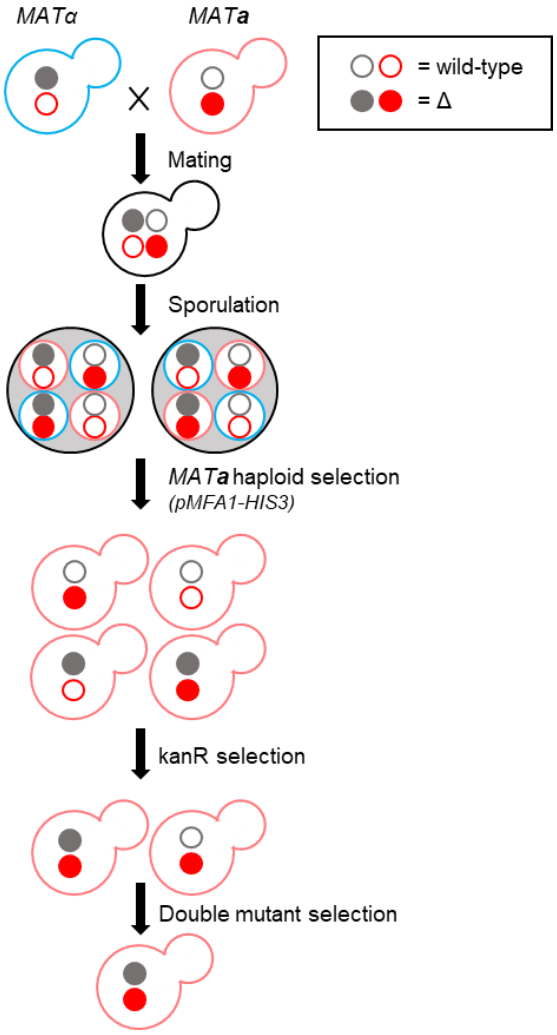


Figure 1. Schematic outline of the Synthetic Genetic Array (SGA) procedure for generation of double mutants by crossing of a query strain to the mutants of the yeast deletion collection. Figure adapted from (74).

The yeast life cycle

In its haploid form, *S. cerevisiae* exists as one of two mating types: either *MAT**a*** or *MATα*. These can grow vegetatively, reproducing by mitosis but two cells of different mating types can also mate and fuse to form a diploid *MAT**a**/MATα* cell. Diploid cells replicate by mitosis and can additionally undergo meiosis via sporulation. Because yeast has a strong preference to being diploid, haploid yeast are also able to switch their mating type, producing a mother and daughter cell of the opposite mating types and thereby enabling self-diploidisation - a mechanism known as homothallism (77). The life cycle of yeast is illustrated in Figure 2 and yeast mating is covered in greater detail below.

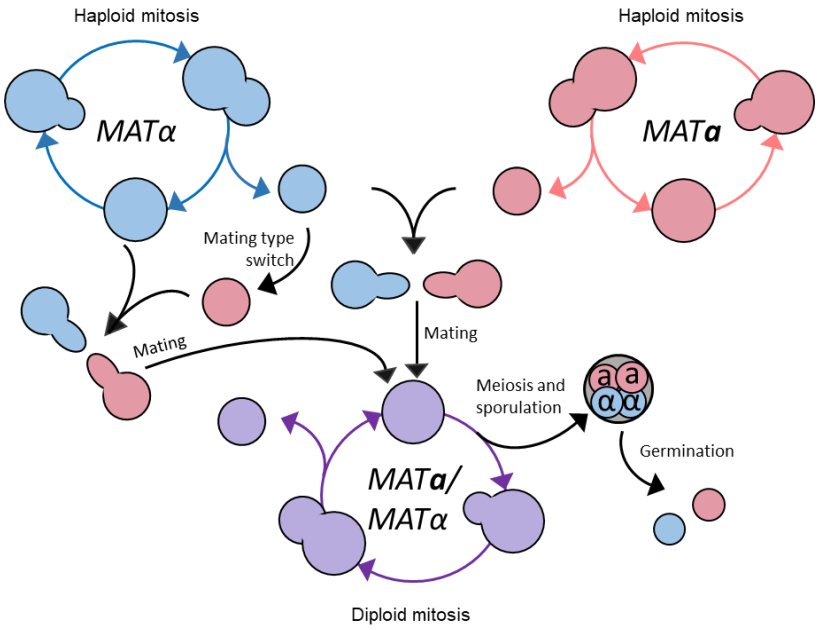


Figure 2. Schematic representation of the yeast life cycle. For simplicity, mating-type switching is only illustrated for a *MATα* cell.

Mating

Yeast mating is mediated by two different pheromones: α-factor and **a**-factor. These are diffusible peptides produced and secreted by *MATα* and *MAT***a** cells, respectively. The two mating types also express genes for two different membrane-bound G protein coupled receptors (GPCRs) – Ste3 in the case of *MATα* and Ste2 in the case of *MAT***a**. Upon binding of pheromones secreted by the opposite mating type, these GPCRs signal to activate physiological responses to prepare the cell for mating (78). The response is mediated by interaction with a guanine nucleotide-binding protein (G protein) associated with the intracellular part of the receptor, and is then relayed by a mitogen activated protein kinase (MAPK) cascade. Two main targets for phosphorylation of this cascade is the Ste12 and Far1 protein. Ste12 is a transcription factor and upon phosphorylation it is known to regulate the expression of more than 200 genes, including genes for both promoting and repressing the mating pathway and genes needed for the fusion of the two mating cells (79). Far1 is a multifunctional regulatory protein, having a principal role in promoting cell cycle arrest in G1 and polarised cell elongation and formation of a mating projection towards the direction of the mating partner (80,81). Upon contact with each other, the plasma membrane of the two mating cells fuse to form a zygote. Then, in the final step of mating, the two nuclei of the zygote combine in a process known as karyogamy, resulting in a diploid cell. The whole process, from pheromone sensing to nuclei fusion, takes approximately four hours (78).

Mating type switching

Under stressful conditions or when suffering starvation, diploid *S. cerevisiae* may undergo meiosis, producing spores that are resistant to the external milieu and which can resume growth (germinate) once conditions improve. Sporulation increases chances of survival under harsh conditions and having a diploid genome thus offers an evolutionary advantage to *S. cerevisiae*, which have evolved the ability to mate and form diploids even in a community of haploid cells of the

same mating type (77). This is possible by mechanisms causing a fraction of the haploid cells to switch mating type - a phenomenon known as homothallism or self-fertilisation. The mating type switch is an intricate process involving many regulatory mechanisms and to understand the processes, the function of the two mating-type alleles *MAT***a** and *MATα*, should be studied in greater detail. The *MAT* locus is positioned in the centre of the right arm of chromosome III in yeast. The locus is composed of five distinct regions referred to as W, X, Y Z1 and Z2. The two mating type alleles *MAT***a** and *MATα* only differ in a region of approximately 700 bp designated Y**a** and Yα, respectively, and encode genes that govern the behaviour of the two different mating types. In addition to the active *MAT* locus, chromosome III also harbours two silent (or cryptic) copies of the *MATa* and *MATα* alleles, which are key in the recombinational process that enables haploid yeast to switch mating type. The loci, termed *HMLα* and *HMR***a**, are positioned at the left and right arm, respectively and their expression is silenced due to the presence of densely packed heterochromatin DNA structures, which formation is mediated by *SIR* gene products (82).The positions of active and silenced *MAT* loci on chromosome III are illustrated in Figure 3.

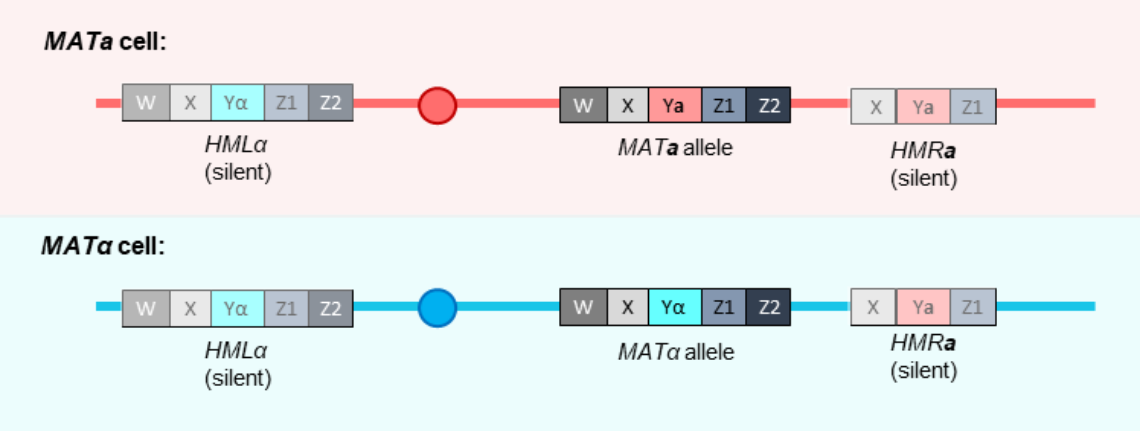


Figure 3. Location and structure of active and silent mating-type cassettes on chromosome III.

In a haploid cell undergoing mating-type switching, replacement of the Y sequence of the active *MAT* locus is triggered by a double-stranded break (DSB) in the region, which is subsequently repaired by homologous recombination using one of the cryptic *MAT* loci as a template (83,84). The DSB is induced by an endonuclease encoded by the *HO* gene. This gene is under tight regulatory control and is only expressed in mother cells for a short time during the G1 phase (85). The HO endonuclease recognises a 24 bp sequence present in both Y**a** and Yα (86). However, although also present in the silent copies, this sequence is not available for cleavage by HO due to nucleosome occupancy of the sites. Thus, upon HO cleavage, DNA repair mechanisms are activated identifying the homologous sequences present in the *HMLα* or *HMR***a** and prompting their use as templates for homologous recombination. Owing to the presence of recombination enhancing elements downstream of the *HMLα* locus, this is the preferred donor for DSB repair in *MAT***a** cells, which express **a**-specific proteins that are likely to contribute to this enhancing effect. In *MATα* cells on the other hand, α-specific proteins suppress the expression of **a**-specific genes and also bind to and suppress the enhancing sequences

downstream the *HML**α***, thereby instead making the *HMR**a*** the preferred substrate for repair in these cells (77). In either scenario, the silent *MAT* locus is used as a template for DNA repair and by gene conversion, the locus replaces the active *MAT* locus thereby causing a switch of mating type. The HO endonuclease is unstable and is quickly degraded so that once the mating locus has been switched, the HO is no longer present to cut again (87).

DNA double-stranded break repair in *S. cerevisiae*

Double strands breaks in the DNA of an organism can arise due to many reasons, including ionising radiation, DNA-damaging chemicals, reactive oxygen species, and stalling replication forks during DNA replication. Furthermore, during meiosis, hundreds of programmed DSBs are introduced across the genome as a means to pair up homologous chromosomes and facilitate exchange of genetic material (88). As these breaks may be lethal if left unrepaired, the cell has evolved sophisticated mechanisms for detecting and repairing them. Two types of evolutionary conserved DNA repair pathways are accessible to the cell: homologous recombination (HR) and non-homologous end joining (NHEJ). These repair mechanisms are actively competing with each other and in yeast, the choice is governed by both ploidy and cell cycle phase (89). In contrast to mammalian cells, the great majority of DSBs in yeast are repaired by HR. This preference of using HR to repair DSBs is one of the contributing factors to the popularity of *S. cerevisiae* as a model eukaryote and biotechnological host, as exogenous DNA is easily integrated in the genome of the organism by the native HR machinery using as little as 40 bp homology (90). The following sections will describe the different mechanisms of DNA repair in greater detail and although focused on *S. cerevisiae*, many of the described pathways are conserved in higher eukaryotes.

Homologous recombination

Upon a DSB, the first step of homologous recombination is initiated by resection of the flanking DNA by 5' to 3' exonucleases, leaving long single-stranded DNA sequences exposed on each side of the break. This resection of ends constitutes the major decision point between HR and NHEJ in the cell (91). Following this initial step, there are three major types of HR available to the cell to repair the break: single-strand annealing, gene conversion and break-induced replication, each of which is described in the following sections and schematically illustrated in Figure 4.

Single strand annealing

If the DSB is flanked by repeating sequences, resection may expose single-stranded ends that can anneal to each other, repairing the DSB but creating a deletion in the process. This repair mechanism, called single-strand annealing (SSA) is the simplest form of HR and can occur with as little as 30 nt of homologous flanking sequences (93)

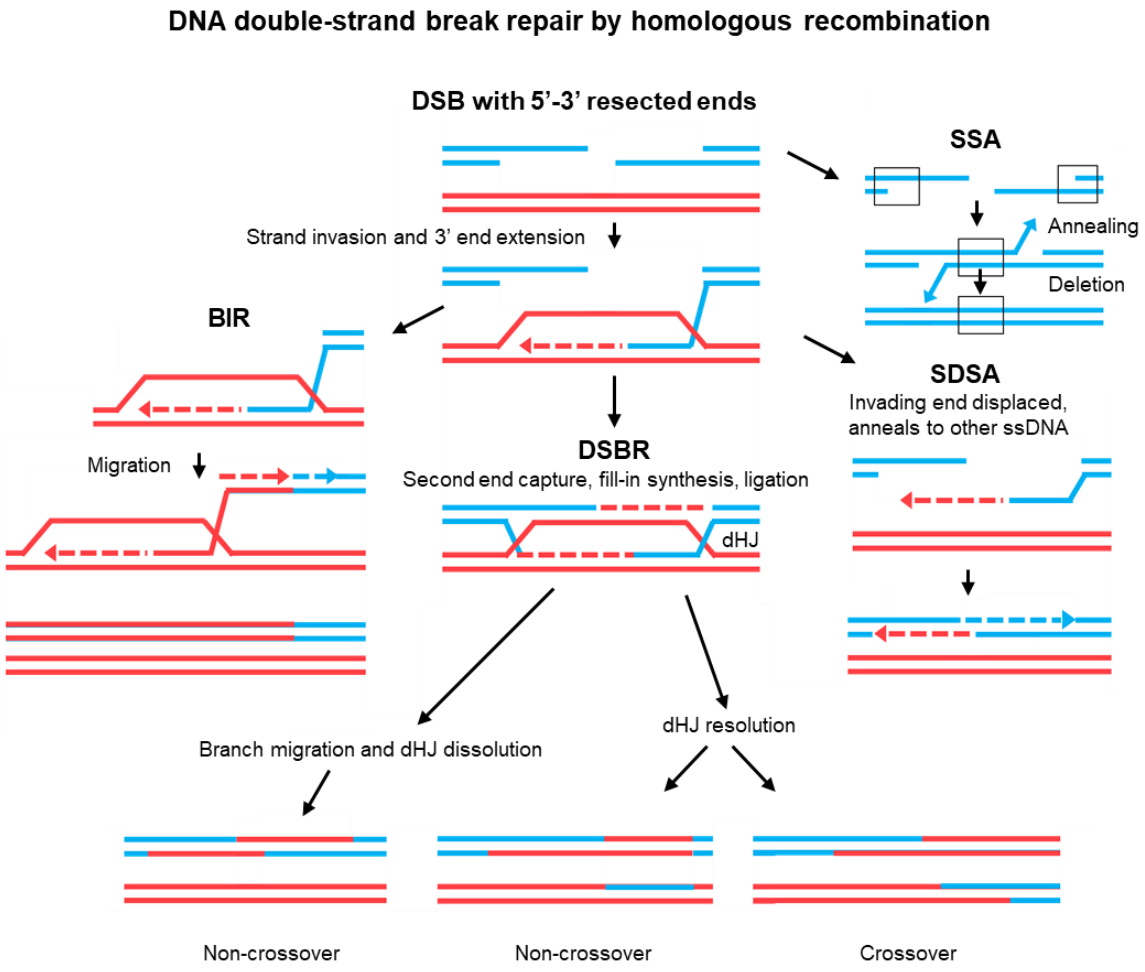


Figure 4. Simplified representation of mechanisms for DNA double strand break (DSB) repair by homologous recombination. SSA = single strand annealing; BIR = break induced replication; DSBR = double-strand break repair; dHJ = double Holliday junction; SDSA = synthesis-dependent strand annealing, Figure adapted from (92).

Gene conversion – with or without crossing over

Gene conversion is distinct from SSA as it depends on the invasion of homologous DNA located elsewhere in the genome by the resected single strand ends of the DSB. This requires the action of several proteins of which only the most central ones are discussed here. Initially, the exposed single-stranded ends are covered by Replication protein A (RPA), protecting them from degradation and preventing the formation of secondary structures. Then, in order to promote strand invasion, Rad52 displaces RPA enabling Rad51 to assemble onto the single-stranded DNA to form a nucleoprotein filament (94). The Rad51-covered ends then engage in an extensive search for homologous sequences throughout the genome, located either on a sister chromatid, a homologous chromosome or at an ectopic location. Once homology is identified, Rad51 promotes strand exchange in which the single-stranded end invades the double-stranded DNA by base pairing with the complementary strand. The displaced strand forms a displacement loop (D-loop) within which DNA polymerase can then assemble and initiate extension of the 3' end of the invading strand (95). In this manner,

information contained in the homologous DNA template is copied into the extending single-strand and can be used to repair the DSB. Two distinct modes of gene conversion-mediated repair exists: synthesis-dependent strand annealing (SDSA) and double-strand break repair (DSBR). Depending on the mode and mechanisms employed during the repair process, gene conversion can either be accompanied by a no crossing-over or a crossing-over event. In the latter case, the broken chromosome and the invaded chromosome, exchange DNA between the point of the DSB and the telomere. If the repair involves the use of a sister chromatid, such an event is typically genetically silent, but if instead a homologous chromosome is used in the repair, this may thus result in loss of heterozygosity.

In the case of SDSA, the newly synthesized strand dissociates from the invaded DNA and is captured by the opposite single-stranded DNA tail of the break. The copied DNA is then used as a primer in a second round of DNA synthesis followed by ligation that repairs the DSB without any associated crossover of template DNA. SDSA is the prevalent repair pathway in mitotic cells (96,97). Alternatively, after strand invasion and DNA synthesis, the second single-stranded end of the DSB may be captured by the D-loop, resulting in the formation of an intermediate double Holliday junction (dHJ). This mode of gene conversion, referred to as DSBR, can have two outcomes, resulting in either crossover or non-crossover depending on how the dHJ is resolved (94). The DSBR pathway appears to be biased towards crossovers (98) and it is the dominant repair pathway in meiotic cells. In these cells, repair of induced DSBs by crossing over is crucial for ensuring proper pairing and segregation of homologous chromosomes and for the genetic scrambling of parental genomes.

Break-induced replication

As for gene conversion, break-induced replication is dependent on the action of Rad51 and the invasion of homologous DNA by the single-stranded ends of the DSB. BIR is employed when only one of the resected ends share enough homology with the template DNA. In this mode of repair, the invading strand may replicate the template DNA until reaching its telomere region or until encountering a replication fork working in the opposite direction. After some delay, as the extending single-stranded DNA gradually dissociates from the moving D-loop, it can act as a template for generating the second strand of copied DNA, thereby repairing the DSB. Replication during BIR is error-prone and mutations frequently occur. Furthermore, the replication bubble in BIR is highly unstable and may even jump between different templates during the replication process (99).

Nonhomologous end-joining

The second type of repair available upon a DSB is nonhomologous end-joining (NHEJ). In contrast to HR, NHEJ is, as the name suggests, independent of homologies between two recombining DNA molecules. The process involves no (or very limited) resection of the double stranded DNA at the break and ultimately the ends are ligated together again either without or with very little base pairing occurring at the location of the break (91). The process is dependent on the Ku protein complex – a highly conserved protein in eukaryotes. The protein, which is a heterodimer, consists of the Ku70 and Ku80 protein and has a central role in binding to the double stranded ends of the DSB, preventing their resection (100) and recruiting factors for NHEJ (101). In mammalian cells, NHEJ is the preferred DNA repair pathway throughout most of the cell cycle, with HR only dominating in mid S phase (102). In *S. cerevisiae* on the other hand, HR is the main

mechanism of DSB repair and NHEJ is only dominant in haploid cells during G1 phase (103–106). NHEJ has traditionally been regarded as an error-prone repair process as it may result in deletion or insertion of nucleotides at the point of repair. In coding DNA sequences this may cause frame-shift mutations deleterious to protein function. More recently though, this view has been increasingly challenged and depending on the nature of the DSB, NHEJ may also provide precise repair (107).

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 technology

As detailed above, homologous recombination is efficient in *S. cerevisiae* and yeast readily incorporates linear exogenous DNA with as little as 40 bp homology to the yeast genome, making gene editing a relatively facile process when compared to other organisms. However, conventional editing based on homologous recombination still requires the use of markers to accompany the desired edit as a means to select for the correct clones. The number of available markers thus introduces a limit for genetic edits and although marker recycling systems such as the Cre-loxP (108) are available, it adds significant time to each editing cycle and may risk chromosomal rearrangements. By introducing a DSB at a chromosomal locus, the frequency of incorporation of linear DNA by homologous recombination can be increased thousand-fold in yeast (109). As the DSB is lethal if left unrepaired it may also function as a selection mechanism, eliminating the need for markers. Introducing DSBs at specified genomic locations has thus been a central aim for improving gene editing. The ability of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas9 to make targeted, precise cuts in the genome by guidance of RNA has therefore revolutionised the field of genetic engineering, allowing genomes to be edited at an unprecedented speed.

In its native form, CRISPR-Cas9 constitutes an adaptive immune system employed by bacteria and archaea to fend off viral infections and invading foreign DNA (Figure 5A). The system has two integral parts: a CRISPR array and its CRISPR associated (Cas) proteins, both encoded in the host genome. The CRISPR array consists of multiple repetitive sequences interspaced by short stretches of variable sequences – so called spacers – originating from prior viral infections. When transcribed and processed by different cas proteins, the CRISPR array produces multiple CRISPR RNAs (crRNAs), each consisting of a variable spacer sequence and a common sequence derived from the CRISPR repeats. The repeat sequence of each crRNA then hybridises with a second RNA molecule, referred to as trans-activating CRISPR RNA (tracrRNA). In this configuration, the hybrid RNA can complex with yet another type of cas protein – an RNA guided endonuclease of which the most well-known is the Cas9 enzyme *of Streptococcus pyogenes*. Interaction with the crRNA-tracrRNA, enables Cas9 to bind to and cut DNA sequences complementary to that of the spacer in the RNA molecule and this DNA is referred to as a protospacer. Thus, upon infection with foreign genetic material of which a spacer sequence has previously been incorporated in the CRISPR array of the host, Cas9 will recognise the protospacer sequence and cut the DNA, thereby eliminating the threat. To avoid the host's own genome being targeted by Cas9, the endonuclease is only able to cut protospacer sequences that are followed by a so called protospacer adjacent motif (PAM). PAM sequences vary between different Cas endonucleases. The most well-studied is the class II enzymes, including *Sp*Cas9 which PAM requirement is 'NGG' and which cuts its targeted DNA 3 nt upstream of this sequence (110). Removal of the PAM sequence completely abolishes the endonuclease activity of

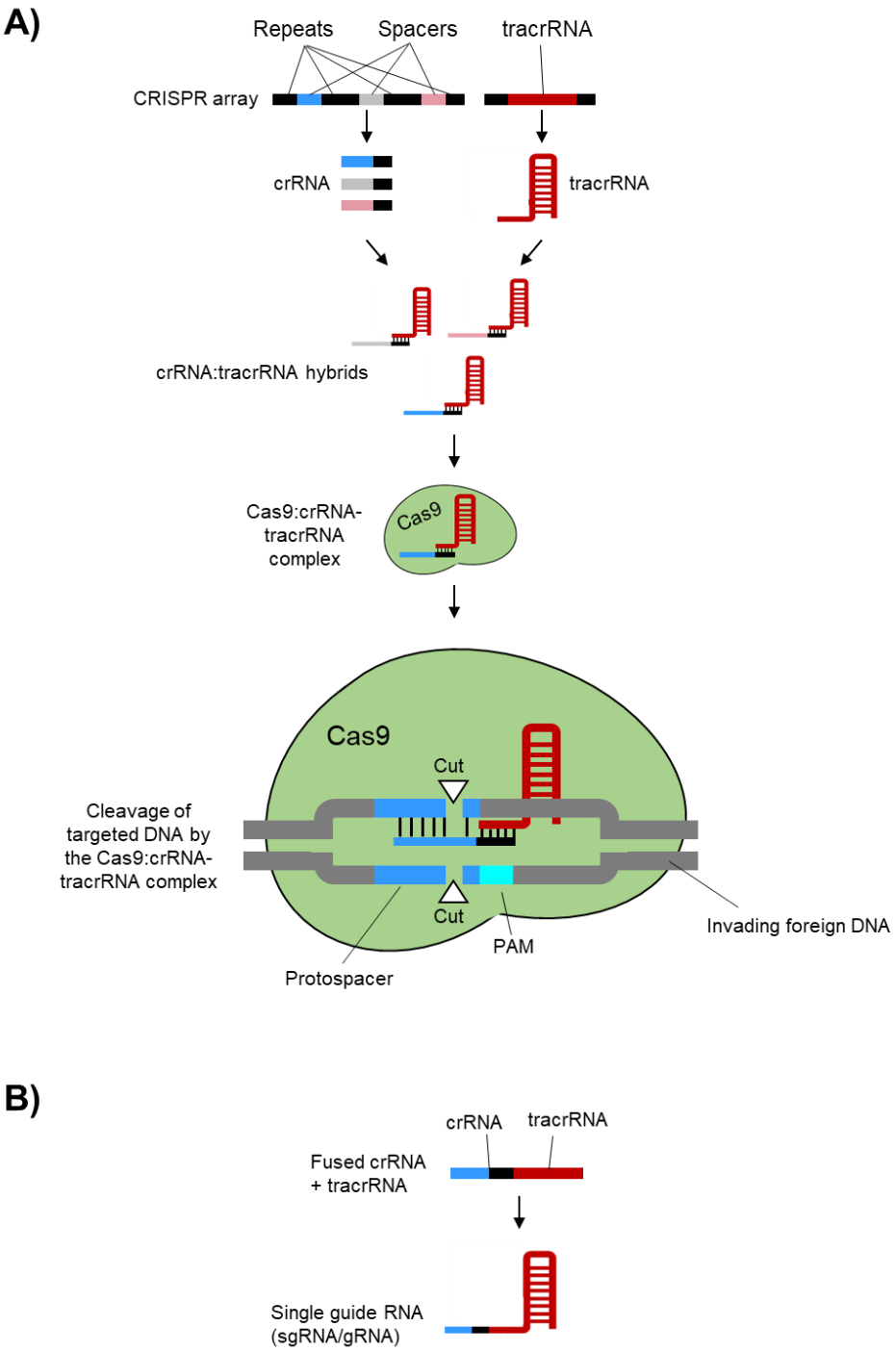


Figure 5. Mechanism of action behind targeted DNA double-strand breaks mediated by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9. (A) Native CRISPR-Cas9 system. (B) Design of a single guide RNA (sgRNA or gRNA) construct used in genetic engineering applications of CRISPR-Cas9.

Cas9. Similarly, mutations to 12th-20th nucleotide of the protospacer (the so called seed sequence) severely effects cutting efficiency (111).

Once the functionality of the CRISPR-Cas9 system was understood and its mechanism of action deduced, its application in RNA-guided genetic engineering was quickly realised and the first use of CRISPR-Cas9 outside its native host was demonstrated in 2011 (112). Shortly thereafter, it was demonstrated that a shortened crRNA of 20 nt is sufficient to guide Cas9 to its genomic target, and that Cas9 can easily be reprogrammed to target a sequence of choice by changing this 20 nt crRNA (110). Furthermore, it was found possible to fuse the crRNA and the tracrRNA into a single functional unit called a single guide RNA (sgRNA or gRNA), containing both the Cas9-binding sequence and the spacer sequence guiding the endonuclease to its target (Figure 6B) (113). In 2013, the first ever use of CRISPR-Cas9 for targeted genome editing in eukaryotes was reported in two different studies performing genetic engineering in human and mouse cells (114,115). CRISPR-Cas9 gene editing has now been adapted for use in a wide range of species including important model organisms such as *Escherichia coli* (111), tobacco plant *Arabidopsis thaliana* (116) and fruit fly *Drosophila melanogaster* (117); and in crops important for human food supply such as rice, wheat (118) and maize (119). CRISPR-Cas9 is also being used to edit the genomes of a large number of yeast species, including *S. cerevisiae* (120), *Kluyveromyces lactis* (121), *Komagataella phaffii* (122), and *Yarrowia liplolytica* (123), and in filamentous fungi such as *Aspergillus niger* and *Aspergillus nidulans* (124).

CRISPR-Cas9 in *S. cerevisiae*

The first use of CRISPR-Cas9 in yeast was demonstrated by Di Carlo *et al.* in 2013. The study made use of a Cas9 gene from *S. pyogenes* codon-optimised for use in human cells and fused with an SV40 nuclear localisation signal at its 3' end. The Cas9 gene was expressed from a centromeric plasmid under control of the strong *TEF1* promoter and was then used in conjunction with a 2µ plasmid from which gRNA was expressed using the RNA polymerase III (pol III) *SNR52* snoRNA promoter, and the 3' sequence flanking the yeast tRNA gene *SUP4* as terminator. As for other orthogonal CRISPR-Cas9 systems, it was crucial to identify these regulatory elements that allowed the transcribed gRNA to remain untranslated and to fold and interact correctly with the Cas9 enzyme. For this reason, promoters of RNA polymerase II (pol II) - the main polymerase for mRNA transcription - cannot directly be used for expression of gRNAs, as transcripts would be processed and transported out of the nucleus for translation. When the gRNA and Cas9 expressing plasmids were combined in the same cell, the study of DiCarlo *et al.* measured a 130-fold increase in HR upon Cas9 cleavage in yeast. Close to 100 % editing efficiency was further observed when co-transforming haploid, Cas9-expressing yeast with a gRNA plasmid targeting the *CAN1* gene and a 90 bp donor DNA for disruption of the locus. The p*SNR52*-t*SUP4* system has since been used in a number of studies performing highly efficient single-target and multiplex editing in haploid and diploid laboratory strains (121,125,126). The gRNA system has also been demonstrated to be functional in diploid and polyploid industrial strains (127–129).

References

1. Steffen W, Persson Å, Deutsch L, Zalasiewicz J, Williams M, Richardson K, *et al.* The anthropocene: From global change to planetary stewardship. Ambio. 2011;40(7):739–61.
2. United Nations. World Population Prospects 2019. Dep Econ Soc Aff Popul Div . 2019;1–25.
3. Tuckett R. Short-Term Energy Outlook (STEO) Forecast highlights. US EIA - Short-Term Energy Outlook. 2018;(April 2017):1–48.
4. Arrhenius S. On the influence of carbonic acid in the air upon the temperature of the ground. London, Edinburgh Dublin Philos Mag J Sci. 1896;41(251).
5. Richardson K, Steffen W, Schellnhuber HJ, Alcamo J, Barker T, Kammen DM, *et al.* Climate Change - Global Risks, Challenges & Decisions: Synthesis Report. København: Museum Tusculanum. 2009;
6. Allen M, Babiker M, Chen Y, Taylor M, Tschakert Australia P, Waisman H, *et al.* IPCC, 2018: Summary for Policymakers. Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change,. 2018.
7. Piotrowski S, Carus M, Carrez D. European Bioeconomy in Figures. Ind Biotechnol. 2016;12(2):78–82.
8. European Commission. Preparing for our future: Developing a common strategy for key enabling technologies in the EU. Com(2009) 512. 2009;1–11.
9. Debergh P, Valentijn B, Els Van de V. Jobs and growth generated by industrial biotechnology in Europe . 2016.
10. Hasan F, Shah AA, Hameed A. Industrial applications of microbial lipases. Enzyme Microb Technol. 2006;39(2):235–51.
11. Wendisch VF. Metabolic engineering advances and prospects for amino acid production. Metab Eng . 2020;58(Sp):17–34.
12. Nielsen J, Larsson C, van Maris A, Pronk J. Metabolic engineering of yeast for production of fuels and chemicals. Curr Opin Biotechnol . 2013;24(3):398–404.
13. Choi SY, Rhie MN, Kim HT, Joo JC, Cho IJ, Son J, *et al.* Metabolic engineering for the synthesis of polyesters: A 100-year journey from polyhydroxyalkanoates to non-natural microbial polyesters. Metab Eng . 2020;58(Sp):47–81.
14. Elander RP. Industrial production of β-lactam antibiotics. Appl Microbiol Biotechnol. 2003;61(5–6):385–92.
15. Nielsen J. Production of biopharmaceutical proteins by yeast. Bioengineered. 2013;4(4):207–11.
16. Moon HG, Jang YS, Cho C, Lee J, Binkley R, Lee SY. One hundred years of clostridial butanol fermentation. FEMS Microbiol Lett. 2016;363(3).
17. Ahn JH, Jang YS, Lee SY. Production of succinic acid by metabolically engineered microorganisms. Curr Opin Biotechnol . 2016;42:54–66.
18. Nakamura CE, Whited GM. Metabolic engineering for the microbial production of 1,3-propanediol. Curr Opin Biotechnol. 2003;14:454–9.
19. de Jong S, Hoefnagels R, Wetterlund E, Pettersson K, Faaij A, Junginger M. Cost optimization of biofuel production – The impact of scale, integration, transport and supply chain configurations. Appl Energy . 2017;195:1055–70.
20. Gustavsson M, Lee SY. Prospects of microbial cell factories developed through systems metabolic engineering. Microb Biotechnol. 2016;9(5):610–7.

21. McGovern PE, Zhang J, Tang J, Zhang Z, Hall GR, Moreau RA, *et al.* Fermented beverages of pre- and proto-historic China. Proc Natl Acad Sci U S A. 2004;101(51):17593–8.

22. Pretorius IS. Synthetic genome engineering forging new frontiers for wine yeast. Crit Rev Biotechnol . 2017;37(1):112–36.

23. Botstein D, Fink GR. Yeast : An Experimental Organism for 21st Century Biology. 2011;189(November):695–704.

24. Siewers V, Mortensen UH, Nielsen J. Genetic Engineering Tools for *Saccharomyces cerevisiae*. In: Baltz R, Demain A, Davies J, Bul A, Junker B, Katz L, *et al.*, editors. Manual of Industrial Microbiology and Biotechnology. Third Edit. American Society for Microbiology; 2010. p. 287–301.

25. Goffeau AA, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, *et al.* Life with 6000 Genes. Science (80-). 1996;274(5287):546, 563–7.

26. Botstein D, Chervitz S, Cherry M. Yeast as a Model Organism. 1997;277(5330):1259–60.

27. Giaever G, Nislow C. The yeast deletion collection: A decade of functional genomics. Genetics. 2014;197(2):451–65.

28. Hong KK, Nielsen J. Metabolic engineering of *Saccharomyces cerevisiae*: A key cell factory platform for future biorefineries. Cell Mol Life Sci. 2012;69(16):2671–90.

29. Medina VG, Almering MJH, Van Maris AJA, Pronk JT. Elimination of glycerol production in anaerobic cultures of a *Saccharomyces cerevisiae* strain engineered to use acetic acid as an electron acceptor. Appl Environ Microbiol. 2010;76(1):190–5.

30. Ito Y, Hirasawa T, Shimizu H. Metabolic engineering of *Saccharomyces cerevisiae* to improve succinic acid production based on metabolic profiling. Biosci Biotechnol Biochem . 2014;78(1):151–9.

31. Maris AJA Van, Geertman JA, Vermeulen A, Groothuizen MK, Winkler AA, Piper MDW, *et al.* Directed Evolution of Pyruvate Decarboxylase-Negative *Saccharomyces cerevisiae*, Yielding a C2-Independent, Glucose-Tolerant, and Pyruvate-Hyperproducing Yeast. Appl Environ Microbiol. 2004;70(1):159–66.

32. Nielsen J, Keasling JD. Engineering Cellular Metabolism. Cell . 2016;164(6):1185–97.

33. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 2003;13(2):244–53.

34. Lu H, Li F, Sánchez BJ, Zhu Z, Li G, Domenzain I, *et al.* A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. Nat Commun. 2019;10(1).

35. Jensen MK, Keasling JD. Recent applications of synthetic biology tools for yeast metabolic engineering. FEMS Yeast Res. 2015;15(1):1–10.

36. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. Nature . 2013;496(7446):528–32.

37. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD. Complete biosynthesis of opioids in yeast. Science (80-) . 2015;349(6252):1095–100.

38. Chen X, Nielsen KF, Borodina I, Kielland-Brandt MC, Karhumaa K. Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. Biotechnol Biofuels. 2011;4(1):21.

39. Yu KO, Jung J, Kim SW, Park CH, Han SO. Synthesis of FAEEs from glycerol in engineered *Saccharomyces cerevisiae* using endogenously produced ethanol by heterologous expression of an unspecific bacterial acyltransferase. Biotechnol Bioeng. 2012;109(1):110–5.

40. Lee W, DaSilva NA. Application of sequential integration for metabolic engineering of 1,2-propanediol production in yeast. Metab Eng. 2006;8(1):58–65.

41. Ishida N, Saitoh S, Tokuhiro K, Nagamori E, Matsuyama T, Kitamoto K, *et al.* Efficient production of L-lactic acid by metabolically engineered *Saccharomyces cerevisiae* with a genome-integrated L-lactate dehydrogenase gene. Appl Environ Microbiol. 2005;71(4):1964–70.

42. Hansen EH, Møller BL, Kock GR, Bünner CM, Kristensen C, Jensen OR, *et al.* De novo biosynthesis of Vanillin in fission yeast (schizosaccharomyces pombe) and baker's yeast (*Saccharomyces cerevisiae*). Appl Environ Microbiol. 2009;75(9):2765–74.

43. Li M, Kildegaard KR, Chen Y, Rodriguez A, Borodina I, Nielsen J. De novo production of resveratrol from glucose or ethanol by engineered *Saccharomyces cerevisiae*. Metab Eng . 2015;32:1–11.

44. Lee SY, Kim HU, Chae TU, Cho JS, Kim JW, Shin JH, *et al.* A comprehensive metabolic map for production of bio-based chemicals. Nat Catal. 2019;2(1):18–33.

45. E4tech, Re-Cord, Wur. From the Sugar Platform to biofuels and biochemicals. Final Rep Eur Comm Dir Energy . 2015;183.

46. Wu G, Yan Q, Jones JA, Tang YJ, Fong SS, Koffas MAG. Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. Trends Biotechnol . 2016;34(8):652–64.

47. Caspeta L, Chen Y, Ghiaci P, Feizi A, Baskov S, Hallström BM, *et al.* Altered sterol composition renders yeast thermotolerant. Science (80-). 2014;346(6205):75–8.

48. Smith J, van Rensburg E, Görgens JF. Simultaneously improving xylose fermentation and tolerance to lignocellulosic inhibitors through evolutionary engineering of recombinant *Saccharomyces cerevisiae* harbouring xylose isomerase. BMC Biotechnol. 2014;14:1–17.

49. Yu T, Zhou YJ, Huang M, Liu Q, Pereira R, David F, *et al.* Reprogramming Yeast Metabolism from Alcoholic Fermentation to Lipogenesis. Cell . 2018;174(6):1549-1558.e14.

50. Zhang J, Jensen MK, Keasling JD. Development of biosensors and their application in metabolic engineering. Curr Opin Chem Biol . 2015;28:1–8.

51. Jones GM, Stalker J, Humphray S, West A, Cox T, Rogers J, *et al.* A systematic library for comprehensive overexpression screens in *Saccharomyces cerevisiae*. Nat Methods. 2008;5(3):239–41.

52. Zhang YX, Perry K, Vinci VA, Powell K, Stemmer WPC, Del Cardayré SB. Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature. 2002;415(6872):644–6.

53. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, *et al.* Programming cells by multiplex genome engineering and accelerated evolution. Nature . 2009;460(7257):894–8.

54. Shi S, Ji H, Siewers V, Nielsen J. Improved production of fatty acids by *Saccharomyces cerevisiae* through screening a cDNA library from the oleaginous yeast Yarrowia lipolytica. FEMS Yeast Res. 2015;16(1):1–10.

55. Mira NP, Palma M, Guerreiro JF, Sá-Correia I. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. Microb Cell Fact . 2010;9(1):79.

56. Endo A, Nakamura T, Ando A, Tokuyasu K, Shima J. Genome-wide screening of the genes required for tolerance to vanillin, which is a potential inhibitor of bioethanol fermentation, in *Saccharomyces cerevisiae*. Biotechnol Biofuels . 2008;1(1):3.

57. Fujita K, Matsuyama A, Kobayashi Y, Iwahashi H. The genome-wide screening of yeast deletion mutants to identify the genes required for tolerance to ethanol and other alcohols. FEMS Yeast Res. 2006;6(5):744–50.

58. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science . 1999;285(5429):901–6.

59. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. Nature. 2002;418(6896):387–91.

60. Pierce SE, Davis RW, Nislow C, Giaever G. Genome-wide analysis of barcoded *Saccharomyces cerevisiae*

gene-deletion mutants in pooled cultures. Nat Protoc. 2007;2(11):2958–74.

61. Smith AM, Durbic T, Kittanakom S, Giaever G, Nislow C. Barcode sequencing for understanding drug-gene interactions. Vol. 910, Methods in Molecular Biology. 2012. 55–69 p.

62. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat Methods. 2008;5(8):711–8.

63. Li Z, Vizeacoumar FJ, Bahr S, Li J, Warringer J, Vizeacoumar FS, *et al.* Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat Biotechnol. 2011;29(4):361–7.

64. Mnaimneh S, Davierwala AP, Haynes J, Moffat J, Peng WT, Zhang W, *et al.* Exploration of essential gene functions via titratable promoter alleles. Cell. 2004;118(1):31–44.

65. Huh, K. W, Falvo, V. J, Gerke, C. L, *et al.* Global analysis of protein localization in budding yeast. Nature. 2003;425(6959):686–91.

66. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, *et al.* Global analysis of protein expression in yeast. Nature. 2003;425(6959):737–41.

67. Gibney PA, Lu C, Caudy AA, Hess DC, Botstein D. Yeast metabolic and signaling genes are required for heat-shock survival and have little overlap with the heat-induced genes. Proc Natl Acad Sci U S A. 2013;110(46).

68. Ryan O, Shapiro RS, Kurat CF, Mayhew D, Baryshnikova A, Chin B, *et al.* Global gene deletion analysis exploring yeast filamentous growth. Science (80-). 2012;337(6100):1352–6.

69. Varela C, Kutyna DR, Solomon MR, Black CA, Borneman A, Henschke PA, *et al.* Evaluation of gene modification strategies for the development of low-alcohol-wine Yeasts. Appl Environ Microbiol. 2012;78(17):6068–77.

70. Kim DU, Hayles J, Kim D, Wood V, Park HO, Won M, *et al.* Analysis of a genome-wide set of gene deletions in the fission yeast Schizosaccharomyces pombe. Nat Biotechnol. 2010;28(6):617–23.

71. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. Mol Syst Biol. 2006;2(1):msb4100050.

72. Noble SM, French S, Kohn LA, Chen V, Johnson AD. Systematic screens of a Candida albicans homozygous deletion library decouple morphogenetic switching and pathogenicity. Nat Genet. 2010;42(7):590–8.

73. Patterson K, Yu J, Landberg J, Chang I, Shavarebi F, Bilanchone V, *et al.* Functional genomics for the oleaginous yeast Yarrowia lipolytica. Metab Eng . 2018;48(March):184–96.

74. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science . 2001;294(5550):2364–8.

75. Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, *et al.* A robust toolkit for functional profiling of the yeast genome. Mol Cell. 2004;16(3):487–96.

76. Suzuki Y, Onge RPS, Mani R, King OD, Heilbut A, Labunskyy VM, *et al.* Knocking out multigene redundancies via cycles of sexual assortment and fluorescence selection. Nat Methods. 2011;8(2):159–64.

77. Haber JE. Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. Genetics. 2012;191(1):33–64.

78. Bardwell L. A walk-through of the yeast mating pheromone response pathway. Peptides. 2004;25(9):1465–76.

79. Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, *et al.* Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. Science (80-). 2000;287(5454):873–80.

80. Chang F, Herskowitz I. Identification of a gene necessary for cell cycle arrest by a negative growth factor of yeast: FAR1 is an inhibitor of a G1 cyclin, CLN2. Cell. 1990;63(5):999–1011.

81. Nern A, Arkowitz RA, Nem A, Arkowitz RA. A Cdc24p-Far1p-Gβγ Protein Complex Required for Yeast Orientation during Mating. J Cell Biol. 1999;144(6):1187–202.

82. Hickman MA, Froyd CA, Rusche LN. Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. Eukaryot Cell. 2011;10(9):1183–92.

83. Strathern JN, Klar AJS, Hicks JB, Abraham JA, Ivy JM, Nasmyth KA, *et al.* Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the MAT locus. Cell. 1982;31(1):183–92.

84. Kostriken R, Strathern JN, Klar AJS, Hicks JB, Heffron F. A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. Cell. 1983;35(1):167–74.

85. Nasmyth K. The Determination of Mother Cell-Specific Mating Type Switching in Yeast by a Specific Regulator of HO Transcription. EMBO J. 1987;6(1):243–8.

86. Nickoloff JA, Chen EY, Heffron F. A 24-base-pair DNA sequence from the MAT locus stimulates intergenic recombination in yeast. Proc Natl Acad Sci U S A. 1986;83(20):7831–5.

87. Kaplun L, Ivantsiv Y, Bakhrat A, Tzirkin R, Baranes K, Shabek N, *et al.* The F-box protein, Ufo1, maintains genome stability by recruiting the yeast mating switch endonuclease, Ho, for rapid proteasome degradation. Isr Med Assoc J. 2006;8(4):246–8.

88. Robert T, Vrielynck N, Mézard C, de Massy B, Grelon M. A new light on the meiotic DSB catalytic complex. Semin Cell Dev Biol . 2016;54:165–76.

89. Haber JE. Partners and pathways - Repairing a double-strand break. Trends Genet. 2000;16(6):259–64.

90. Lorenz MC, Muir RS, Lim E, McElver J, Weber SC, Heitman J. Gene disruption with PCR products in *Saccharomyces cerevisiae*. Gene. 1995;158(1):113–7.

91. Symington LS, Gautier J. Double-Strand Break End Resection and Repair Pathway Choice. Annu Rev Genet. 2011;45(1):247–71.

92. Symington LS, Rothstein R, Lisby M. Mechanisms and regulation of mitotic recombination in *Saccharomyces cerevisiae*. Genetics. 2014;198(3):795–835.

93. Sugawara N, Ira G, Haber JE. DNA Length Dependence of the Single-Strand Annealing Pathway and the Role of *Saccharomyces cerevisiae* RAD59 in Double-Strand Break Repair. Mol Cell Biol. 2000;20(14):5300–9.

94. Mortensen UH, Lisby M, Rothstein R. Rad52. Curr Biol. 2009;19(16):R676–7.

95. Haber JE. DNA Repair : The Search for Homology. 2018;1700229:1–12.

96. Nassif N, Penney J, Pal S, Engels WR, Gloor GB. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol Cell Biol. 1994;

97. Ferguson DO, Holloman WK. Recombinational repair of gaps in DNA is asymmetric in Ustilago maydis and can be explained by a migrating D-loop model. Proc Natl Acad Sci U S A. 1996;93(11):5419–24.

98. Allers T, Lichten M. Differential timing and control of noncrossover and crossover recombination during meiosis. Cell. 2001;106(1):47–57.

99. Jasin M, Haber JE. The democratization of gene editing : Insights from site-specific cleavage and double-strand break repair. DNA Repair (Amst) . 2016;44:6–16.

100. Clerici M, Mantiero D, Guerini I, Lucchini G, Longhese MP. The Yku70-Yku80 complex contributes to regulate double-strand break processing and checkpoint activation during the cell cycle. EMBO Rep. 2008;9(8):810–8.

101. Zhang Y, Hefferin ML, Chen L, Shim EY, Tseng HM, Kwon Y, *et al.* Role of Dnl4-Lif1 in nonhomologous end-joining repair complex assembly and suppression of homologous recombination. Nat Struct Mol Biol. 2007;14(7):639–46.

102. Karanam K, Kafri R, Loewer A, Lahav G. Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. Mol Cell. 2012;47(2):320–9.

103. Ira G, Pellicioll A, Balijja A, Wang X, Florani S, Carotenuto W, *et al.* DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. Nature. 2004;431(7011):1011–7.

104. Aylon Y, Kupiec M. Cell cycle-dependent regulation of Double-Strand Break repair: A role for the CDK. Cell Cycle. 2005;4(2):259–61.

105. Lisby M, Antúnez de Mayolo A, Mortensen UH, Rothstein R. Cell cycle-regulated centers of DNA double-strand break repair. Cell Cycle. 2003;2(5):479–83.

106. Lisby M, Barlow JH, Burgess RC, Rothstein R. Choreography of the DNA damage response: Spatiotemporal relationships among checkpoint and repair proteins. Cell. 2004;118(6):699–713.

107. Bétermier M, Bertrand P, Lopez BS. Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? PLoS Genet. 2014;10(1).

108. Güldener U, Heck S, Fiedler T, Beinhauer J, Hegemann JH. A new efficient gene disruption cassette for repeateduse in budding yeast. Nucleic Acids Res . 1996;24(13):2519–24.

109. Storici F, Durham CL, Gordenin DA, Resnick MA. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. Proc Natl Acad Sci U S A. 2003;100(25):14994–9.

110. Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A. 2012;109(39):2579–86.

111. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol . 2013;31(3):233–9.

112. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. 2011;39(21):9275–82.

113. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. 2012;337(August):816–22.

114. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, *et al.* RNA-guided human genome engineering via Cas9. Science (80-). 2013;339(6121):823–6.

115. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, *et al.* Multiplex genome engineering using CRISPR/Cas systems. Science (80-). 2013;339(6121):819–23.

116. Li J-F, Norville JE, Aach J, McCormack M, Zhang D, Bush J, *et al.* Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol. 2013;31(8):688–91.

117. Yu Z, Ren M, Wang Z, Zhang B, Rong YS, Jiao R, *et al.* Highly efficient genome modifications mediated by CRISPR/Cas9 in Drosophila. Genetics. 2013;195(1):289–91.

118. Shan Q, Wang Y, Li J, Zhang L, Chen K, Liang Z, *et al.* Targeted genome modification of crop plants using a CRISPR-Cas system. Nat Biotechnol. 2013;31(8):686–8.

119. Liang Z, Zhang K, Chen K, Gao C. Targeted mutagenesis in Zea mays using TALENs and the CRISPR/Cas system. J Genet Genomics . 2014;41(2):63–8.

120. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. 2013;41(7):4336–43.

121. Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, *et al.* Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell Syst . 2015;1(1):88–96.

122. Weninger A, Hatzl A-M, Schmid C, Vogl T, Glieder A. Weninger 2016 - Combinatorial optimization of CRISPRCas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris.. J Biotechnol. 2016;235:139–49.

123. Schwartz CM, Hussain MS, Blenner M, Wheeldon I. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. ACS Synth Biol. 2016;5(4):356–9.

124. Nødvig CS, Nielsen JB, Kogle ME, Mortensen UH. A CRISPR-Cas9 system for genetic engineering of filamentous fungi. PLoS One. 2015;10(7):1–18.

125. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NGA, van den Broek M, *et al.* CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res. 2015;15(2):1–15.

126. Jakočinas T, Bonde I, Herrgård M, Harrison SJ, Kristensen M, Pedersen LE, *et al.* Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. Metab Eng. 2015;28:213–22.

127. Stovicek V, Borodina I, Forster J. CRISPR-Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae* strains. Metab Eng Commun . 2015;2:13–22.

128. Generoso WC, Gottardi M, Oreb M, Boles E. Simplified CRISPR-Cas genome editing for *Saccharomyces cerevisiae*. J Microbiol Methods . 2016;127:203–5.

129. Zhang GC, Kong II, Kim H, Liu JJ, Cate JHD, Jin YS. Construction of a quadruple auxotrophic mutant of an industrial polyploid *Saccharomyces cerevisiae* strain by using RNA-guided Cas9 nuclease. Appl Environ Microbiol. 2014;80(24):7694–701.

CHAPTER 2: **Development of the CRISPR-Cas9-Selective Ploidy Ablation (CRI-SPA) method for high-throughput, automated manipulation of the yeast deletion collection**

Abstract

The yeast genome-wide deletion collection is a valuable resource for studying and understanding the function of genes in *Saccharomyces cerevisiae* and may also aid the study of corresponding genes in higher eukaryotes. Importantly, by introducing a second loss-of-function (LOF) mutation to the approximately 4800 strains of the deletion collection, genetic interactions can be studied and gene network structures uncovered. To do this, high-throughput methods for manipulation of the collection mutant strains are needed. In this study, we present the development of CRI-SPA (CRISPR-Cas9 Selective Ploidy Ablation) – a mating-based, CRISPR-Cas9 assisted method for high-throughput, automated manipulation of the yeast deletion collection. The CRI-SPA method is independent of meiosis and requires only a short pin replica plating protocol and less than a week to complete the transfer of genetic material to the recipient strain library. We use the CRI-SPA method to efficiently transfer the *ade2Δ* gene deletion to a subset of the yeast deletion collection and we prove its usability in genetic interaction screens by showing the accuracy of the method in detecting the known genetic interaction between the *ADE2* and *ADE3* gene. Incorporation of the CRISPR-Cas9 technology in our method offers great versatility and this study sets the stage for the CRI-SPA method to be used for other types of applications, transferring different types of DNA into diverse strain libraries.

Introduction

Saccharomyces cerevisiae was the first eukaryotic organism to have its genome fully sequenced and has for a long time served as a model eukaryote in physiological and genetic studies. Approximately 60% of yeast genes share significant homology with human genes (1) and it has been estimated that ca. 30% of human genes that are linked to different diseases have a yeast ortholog (2). Fundamental research in yeast has thus frequently uncovered functions of corresponding human genes and discoveries from yeast genetic studies have seen numerous implications in medicine. A vital success factor for these genetic studies was the generation of the genome-wide yeast deletion collection in the late 1990's, in which each non-essential open reading frame (ORF) was systematically deleted in a common haploid strain background (3,4). Since its creation, the deletion collection has been used in countless of genetic studies, investigating biological functions of genes, responses to different types of stress and toxins, and the mechanism of action of drugs (5). In 2015, the collection was even sent to the International Space Station to study genes important for growth and survival under microgravity conditions (6).

In addition to serving as a tool for understanding the function of individual genes, the yeast deletion collection can also be used for genetic interaction studies, where two loss-of-function (LOF) mutations are combined in the same cell. These types of studies were largely enabled through the Synthetic Genetic Arrays (SGA) analysis method developed in 2001 (7). In brief, this method makes use of yeast mating and mating-type specific reporter genes to transfer a mutant allele from a query strain to the strains of the deletion library. The method can be carried out in high-throughput (HT) using

arrayed library strains on rectangular agar plates, and an automated pin replication procedure for transferring strains between different selective media. When studying the phenotype of the resulting double mutants, suppressing or enhancing interactions can be identified, which in turn can reveal structures of genetic networks and metabolic pathways and for this, SGA has proven to be a truly powerful tool (8,9). The method is, however, dependent on meiosis and sporulation to haploidise mated strains, which has a number of drawbacks. Not only does this add significant time to the SGA protocol, which takes around 20 days to complete, it also means that simultaneous transfer of multiple features into the deletion library becomes difficult as modified chromosomes need to co-segregate during meiosis in order to yield a mutant carrying all the intended modifications. With increasing number of modifications spread across multiple chromosomes, the probability of co-segregation decreases (0.5^x, with x being the number of mutation-carrying chromosomes) making the generation and identification of correct spores increasingly difficult and necessitating the use of selectable markers for each modification. Furthermore, sporulation itself may be problematic as the S288C genetic background of the mutants strains is known to have poor sporulation efficiency (only around 12 % (10)). Finally, because of the dependency on meiosis, it is only meaningful to use the SGA method with recipient strains that are isogenic to the query strain, as homologous chromosomes are thoroughly scrambled due to extensive and unavoidable homologous recombination during meiosis.

In 2011, Reid *et al.* presented an alternative mating-based method for HT manipulation of the yeast deletion collection, in which haploidisation of diploids to generate the original strains occur independent of meiosis (11). The method, outlined in Figure 1, is termed Selective Ploidy Ablation (SPA) and makes use of a Universal Donor Strain (UDS) that is specifically designed for its purposes. The strain carries a set of 16 conditionally stable and counter-selectable chromosomes - a dual feature attributable to the integration of a strong *GAL1* promoter and a *Kluyveromyces lactis URA3* marker at the centromere of each chromosome (Figure 1A). When the strain is grown on galactose, expression is induced across the centromere and polymerases are recruited to the site. This abolishes the function of the centromere (12,13) (which is only around 125 bp long in yeast) and it can no longer serve as an assembly point for the kinetochore to which the mitotic spindle attach during cell division (14). Hence, when mated to a recipient strain and grown under inducing conditions, chromosomes originating from the UDS can be selectively lost (ablated) as they lose their ability to propagate correctly into daughter cells (Figure 1B). What is more, complete loss of genetic material from the donor can also be guaranteed by subsequent counter-selection on 5-fluoroorotic acid (5-FOA) containing medium, as the compound is toxic to *URA3* cells (15). The method hence requires recipient strains to be $ura3^{-}$, as in the case of the yeast deletion collection mutants. Similar to the SGA method, the SPA method can be used with recipients strains arrayed on rectangular high-throughput screening (HTS) agar plates, which are then transferred using an automated pinning procedure to different selective plates and in this manner, Reid *et al.* successfully used the SPA method to transfer a plasmid from a Universal Donor Strain into the full deletion library (Figure 1C).

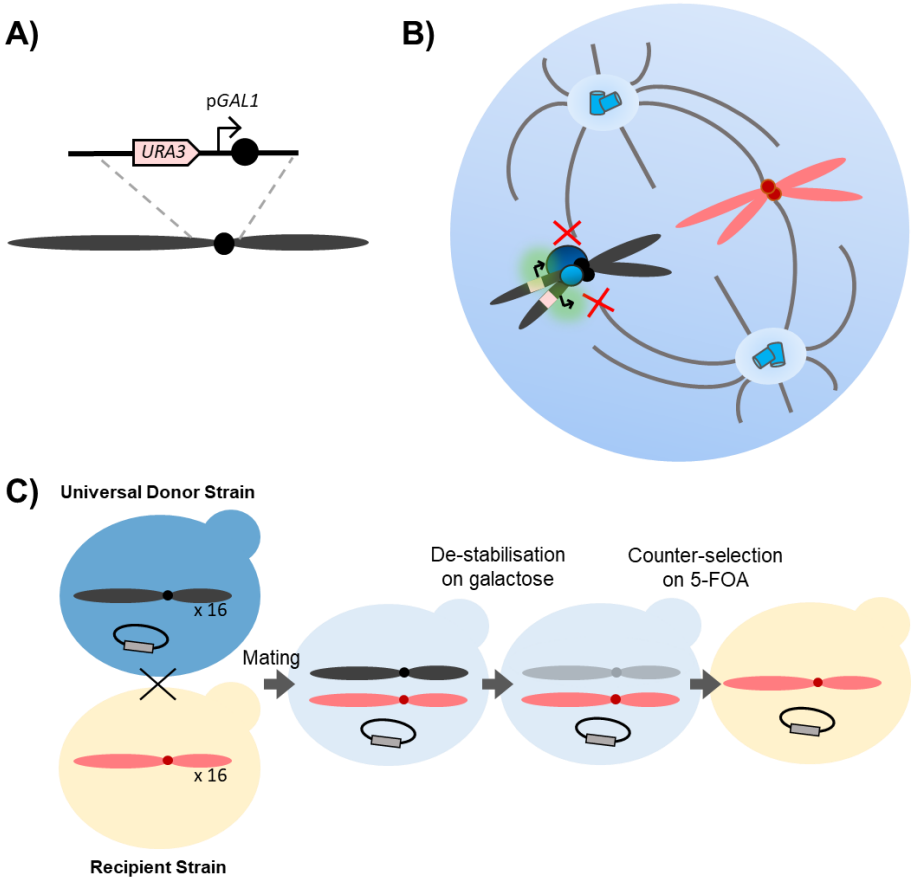


Figure 6. Principle behind the Selective Ploidy Ablation (SPA) method. (A) Chromosomal architecture of the Universal Donor Strain (UDS). The UDS carries 16 chromosomes which have been rendered conditionally stable and counter-selectable owing to the insertion of a *Kluyveromyces lactis URA3* marker and a galactose-inducible p*GAL1* promoter just next to the centromere of each chromosome. (B) Schematic outline of the ablation of UDS chromosome during mitosis. When mated to a recipient strain the chromosomes originating from the UDS can be selectively lost (ablated) by growth on galactose. As transcription is induced across the centromere, transcription machinery proteins are recruited to the site and in a mitotic cell this effectively prevents the assembly of the kinetochore and subsequent anchoring of the mitotic spindle. Hence, as the cell divides, UDS chromosomes are not able to segregate properly into daughter cells and as mitosis continues, cells without any remaining UDS chromosomes arise. (C) SPA procedure for mating-based plasmid transfer. Using the SPA method, plasmids can be transferred to recipient strains by mating to a UDS harbouring the plasmid. After mating, diploids are grown on galactose to destabilise the donor genome while maintaining chromosomes from the recipient. Then, to ensure complete loss of UDS chromosomes, counter-selection against *URA3* is applied by growth on 5-FOA, leaving only the recipient background, now carrying the plasmid.

Since the development of the SPA method, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 technology made its entry into the scientific world, revolutionising the field of genetic engineering in terms of both speed and precision. In this study, we take the method of Reid *et al.* (2011) one step further by combining CRISPR-Cas9 with the SPA technology, developing a novel method that we call CRI-SPA. The use of CRISPR-Cas9 in this method opens up for new possibilities of transferring not only plasmids in high throughput but to also chromosomally integrate transferred DNA in the recipient strains of the deletion collection. This alleviates the issue of plasmid copy number variation and size limitations in terms of the capacity of the plasmid to harbour large DNA constructs – an important issue if considering transfer of complex biochemical pathways. Furthermore, it also allows the method to, in similarity to SGA,

be used for genetic interaction studies in which LOF deletions are transferred to the deletion library. Finally, the use of CRISPR-Cas9 allows marker-free and multiplex transfer of several genetic features simultaneously. Here, we outline the development of the CRI-SPA method and in a proof-of-principle study, we show that the method is functional and can be used to study genetic interactions in a high-throughput manner by transfer of the *ade2Δ* deletion to a subset of the yeast deletion collection. Our work demonstrates the efficiency and potential of the method, which we envision will have useful applications also in metabolic engineering studies as the method can be used to transfer any type of genetic feature, including biosynthetic pathways, to the mutant strains of the deletion collection.

Results

Developing the CRI-SPA concept

Here, the concept of the CRI-SPA method is described and its procedure is also schematically explained in Figure 2. The UDS of CRI-SPA carries a chromosomally integrated copy of the Cas9 gene. Moreover, a *Kl URA3* –p*GAL1* cassette is integrated by the centromere of all 16 chromosomes, allowing their selective ablation and counter-selection. To set the stage for the transfer of a genetic feature to the strains of the deletion library, the genetic feature, flanked by an additional *Kl URA3* marker, is first integrated in the genome of the UDS by Cas9 targeting. The step is mediated by a selectable 2 µm plasmid encoding the relevant gRNA. Upon integration of the feature, Cas9 is no longer able to cut the chromosome of the UDS. Specific gene transfer from the modified UDS to a *ura3* recipient strain containing another selectable feature, e.g. the *KanMX* cassette in the gene deletion library strains, is then performed by the following steps.

1) The UDS carrying the genetic feature of interest and the relevant gRNA plasmid is mated to the recipient strains.
2) Diploids are selected on solid medium via the marker on the gRNA plasmid and the marker in the recipient strains. In the diploid, the Cas9-gRNA complex cleaves the target sequence in the recipient genome. The resulting DNA double-strand break (DSB) is repaired by homologous recombination (HR) using the region containing the genetic feature of interest in the donor chromosome as repair template. In most cases, HR mediated repair is expected to yield a gene conversion event, where the genetic feature of interest is transferred form the donor to the recipient chromosome. However, in some cases, gene conversion may be accompanied by a cross-over event, or alternatively, HR repair may be performed by break induced replication (BIR). Strains resulting from the two latter repair events are undesirable as the repaired chromosome, in addition to the genetic feature of interest, will contain the donor chromosomal sequence contained between the position of the DSB and the telomere (Figure 3). However, due to the presence of the second *URA3* marker flanking the genetic feature of interest, these events are counter-selected during growth on 5-FOA in step 4 below.
3) Cells are next haploidised via SPA by transfer onto galactose, which selectively de-stabilises UDS chromosomes.
4) Transfer to solid medium containing 5-FOA ensures that the *URA3*-marked donor chromosomes have been completely lost and that the recipient strain background is restored in exception for the targeted locus, which now harbours the transferred DNA. Moreover, this final step counter-selects clones in which gene conversion was accompanied by crossover or in which the DSB was repair via BIR.
5) Finally, transformed cells are selected on solid medium containing G418 and are now ready for further analysis.

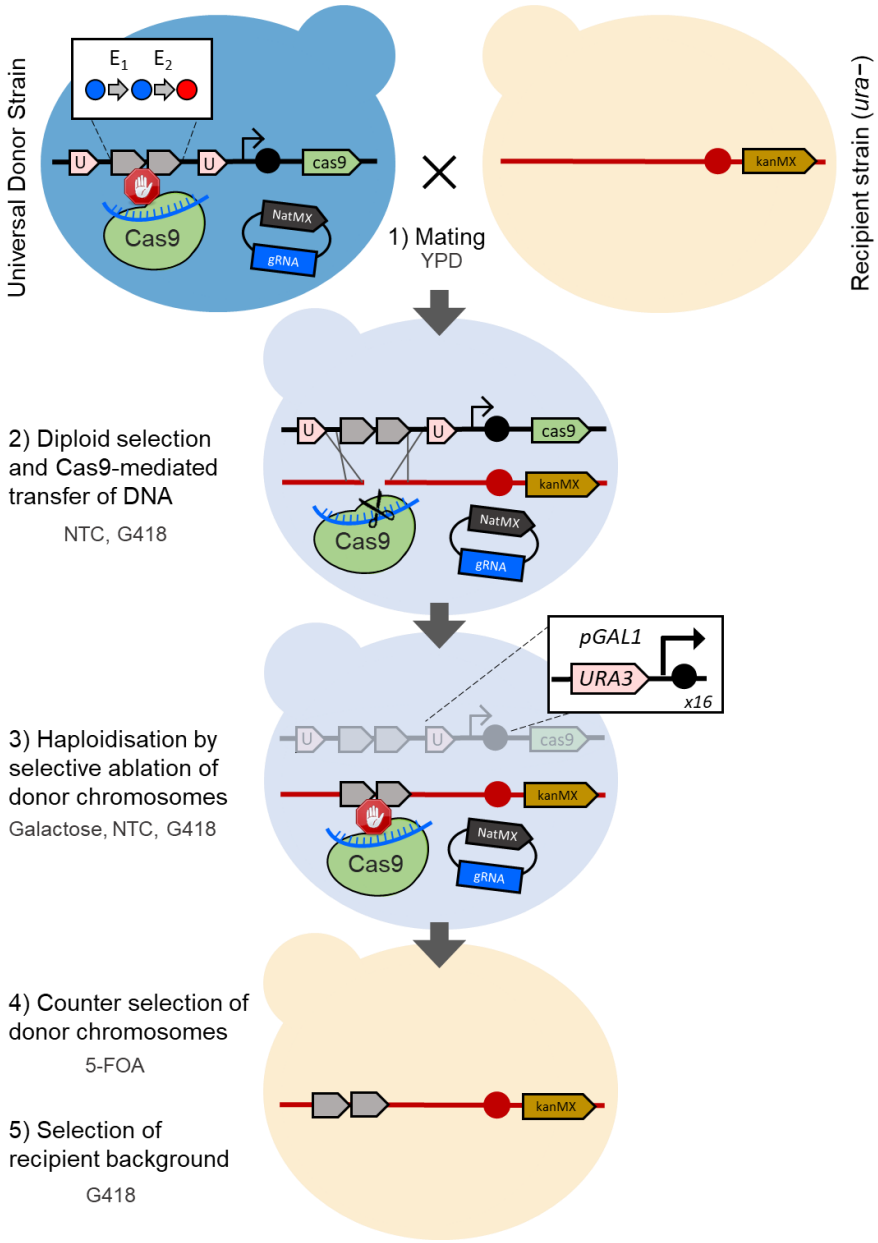


Figure 7. Schematic outline of the CRI-SPA procedure. The Universal Donor Strain (UDS) of the CRI-SPA method expresses a Cas9 gene from a chromosomal integration site and harbours a chromosomally integrated genetic feature – here illustrated with a biosynthetic pathway - which one wishes to transfer in high-throughput to a library of recipient strains. The UDS also carries a gRNA plasmid for targeting the same site at which the genetic feature is integrated, leaving the Cas9-gRNA complex unable to cleave the DNA. However, when mated to a recipient strain, Cas9 can cleave the un-altered chromosome of the recipient, prompting the break to be repaired using the homologous chromosome of the UDS as template. By homologous recombination, the genetic feature of the donor is transferred into the recipient genome. The CRI-SPA UDS also harbours 16 conditionally stable and counter-selectable chromosomes owing to the integration of a *URA3-pGAL1* cassette at the centromere of each chromosome and in a next step, diploid cells are haploidised by growing on galactose causing selective loss of UDS chromosomes. Complete haploidisation is ensured by growing on 5-FOA, counter-selecting any remaining *URA3*-tagged chromosomes. Moreover, as the genetic feature to be transferred is flanked by not only the centromere-adjacent *URA3* marker in the UDS but also a second *URA3* marker upstream of the site, counter-selection on 5-FOA also ensures that no cells with large chromosomal crossovers survive the process.

As for the original SPA method, CRI-SPA can be used in a high-throughput manner, using recipient strains arrayed on HTS agar plates and an automatic pin replication procedure for transferring cells between different selective media. In the following sections, the construction of new CRI-SPA Universal Donor Strains is described and the steps involved in developing an efficient plating protocol for CRI-SPA-mediated manipulation of the yeast deletion library are outlined. Furthermore, the use of the developed CRI-SPA protocol in genetic interaction studies is demonstrated.

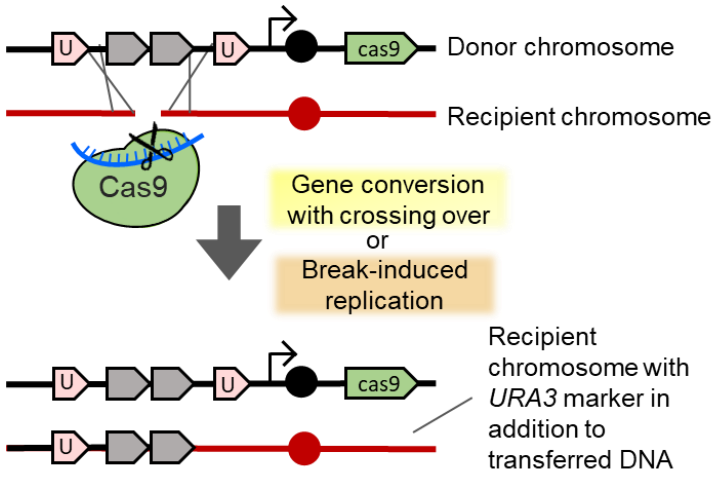


Figure 8. Strategy for counter-selection of recipient chromosomes harbouring genetic material from the donor. The DNA double-stranded break (DSB) in the chromosome of the recipient is repaired by homologous recombination using the homologous chromosome of the Universal Donor Strain as template. In the process, the genetic feature of interest contained within the donor chromosome is expected to be transferred into the recipient chromosome by gene conversion. On occasion, gene conversion is accompanied by crossing-over or, alternatively, the DSB may also be repaired by break-induced replication (BIR). In these scenarios, genetic information of the donor chromosome arm contained between the homology region and the telomere is transferred into the chromosome being repaired. In CRI-SPA, these events can be counter-selected as the DNA to be transferred is flanked by URA3 markers on each side in the donor chromosome. Crossing over or BIR thus results in a recipient cell not only harbouring the intended transferred DNA but also a URA3 marker and can thereby be counter-selected by growth on 5-FOA.

Constructing new Universal Donor Strains for CRI-SPA

To incorporate CRISPR-Cas9 technology in the SPA method, two new Universal Donor Strains were constructed. A *Streptococcus pyogenes* Cas9 gene codon-optimised for use in humans was genomically integrated in the two strains W8164-2B (*MATα* UDS) and W8164-2C (*MATa* UDS), yielding new Universal Donor Strains *Sc*_HO10 and *Sc*_HO11, respectively. The functionality of the Cas9 endonuclease was verified by co-transformation with gRNA plasmid pCfB2311, targeting the *ADE2* locus, and a 90 bp double stranded repair substrate for deletion of the *ADE2* coding sequence (Figure 4A). A control transformation with an empty gRNA plasmid lacking a target sequence was also included. In yeast, *ade2* mutants display a red phenotype (16) and the number of red transformants was used to verify functionality of the Cas9 nuclease. Both *Sc*_HO10 and *Sc*_HO11 produced > 95 % red colonies when transformed with pCfB2311 and the repair substrate (Figure 4B), whereas transformation with the empty control plasmid and repair substrate only yielded white colonies (data not shown).

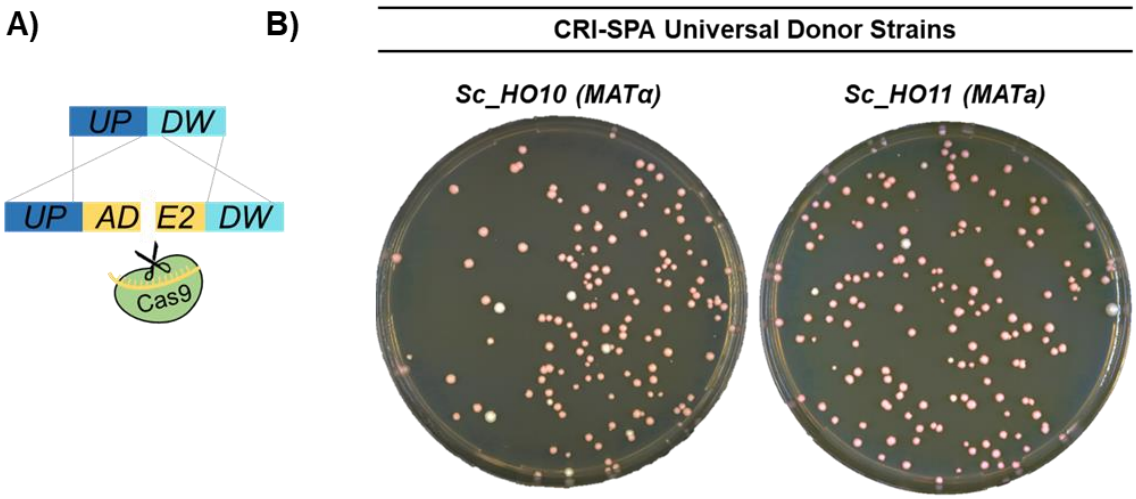


Figure 4. Verifying Cas9 efficiency in the CRI-SPA Universal Donor Strains *Sc*_HO10 and *Sc*_HO11. (A) Experimental set-up, using CRISPR-Cas9 and gRNA to delete the *ADE2* locus by co-transformation with a 90 bp deletion substrate. (B) *Sc*_HO10 and *Sc*_HO11 transformed with the *ADE2* targeting gRNA plasmid pCfB2311 and the 90 bp repair substrate. The red phenotype of colonies indicate successful deletion of the locus.

Developing a CRI-SPA plating protocol

The functionality and efficiency of the CRI-SPA method was next tested, applying for the first time, this method to transfer a gene deletion to a subset of the yeast deletion collection in order to study genetic interactions. Deletion of the *ADE2* gene was chosen as a proof-of-principle test case. The metabolism of adenine synthesis in yeast is outlined in Figure 5. In the synthesis pathway, *ADE2* encodes an enzyme catalysing the conversion of 5'-phosphoribosylaminoimidazole (AIR) into 5'-phosphoribosylaminoimidazole carboxylate (CAIR). This compound is then further converted to adenine by multiple enzymes downstream in the pathway. Wild-type yeast cells have a white/beige appearance (Figure 5A), but if *ADE2* is deleted or otherwise impaired, the intermediate AIR accumulates. Through oxidation and polymerisation of AIR, a red pigment is formed that causes *ade2*Δ cells to appear red (Figure 5B). However, genetic interactions may suppress the red phenotype of *ade2* mutants. For example, Ade3 acts upstream to Ade2 in the synthesis pathway, where it catalyses the conversion of 5,10-tetrahydrofolate (5,10-THF) into 10-formyl-THF, which is further used as a co-factor in downstream catalytic steps. An *ade2 ade3* double mutant therefore does not accumulate AIR and its colonies have a wild-type white/beige appearance (17) (Figure 5C). Thus, using *ADE2* in a proof-of-principle CRI-SPA-mediated transfer of gene deletions, provided not only a visible phenotype for assessing correct transfer but also a simple test system for identification of genetic interactions.

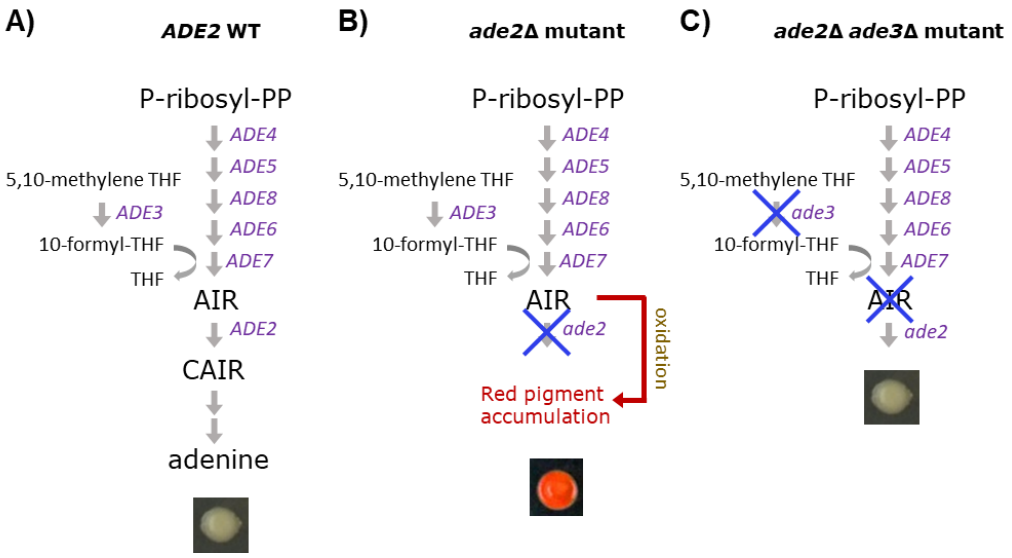


Figure 5. **Synthesis of adenine in yeast.** (A) In a wild-type cell, a number of *ADE* genes catalyse the conversion of phosphoribosyl pyrophosphate (P-ribosyl-PP) to adenine. In this chain of reactions, Ade2 is responsible for the catalysis of the conversion of 5'-phosphoribosylaminoimidazole (AIR) into 5'-phosphoribosylaminoimidazole carboxylate (CAIR). Cells with a functional adenine synthesis pathway display a white phenotype. (B) In an *ade2* mutant cell, conversion of AIR to CAIR is blocked. This causes build-up of AIR, which oxidises and by polymerisation forms a red pigment. Cells with a non-functional *ADE2* therefore displays a red phenotype. (C) Ade3 acts upstream of Ade2 in the synthesis pathway. Mutation of the *ADE3* gene therefore suppresses the red phenotype of an *ade2Δ* mutant and the *ade2*Δ *ade3*Δ double mutant has a white appearance.

The *ADE2* gene was deleted in the Universal Donor Strain *Sc*_HO10 by simultaneous integration of the *hphNT1* marker conferring resistance to hygromycin. Furthermore, an additional *URA3* marker was then inserted on chromosome XV in an intergenic region between the*THI72* gene and the long terminal repeat sequence *YORCdelta18*, approximately 140 kb downstream of the *ADE2* locus and distal to the centromere. This yielded strain *Sc*_HO19 in which the genetic material that was to be transferred to the recipient strains – in this case the *ade2*Δ::*hphNT1* allele – was thus located between two *URA3* markers: one proximal and one distal to the centromere. As explained, this design allowed counter-selection of recipient clones in which gene conversion was accompanied by a crossover or in which HR was mediated through BIR. Finally, *Sc*_HO19 was transformed with gRNA plasmid pHO24, targeting the wild-type *ADE2* locus but not the disrupted allele of the donor.

To assess the functionality of CRI-SPA, the *ade2Δ::hphNT1* allele was then transferred to a subset of the yeast deletion library that harboured the *ade3*Δ mutant and which consisted of the 376 strains contained on plate Δp9 (Supplementary Figure 1). As outlined above, *ade3*Δ is a synthetic suppressor of the *ade2* phenotype and thus functioned as a positive control for detection of genetic interactions. The experiment is summarised in Figure 6. The library strains (*ura3 orfΔ::kanMX*) were first arrayed in a 16 by 24 grid (384 density format) on an HTS agar plate and were then pinned on top of a lawn of the Universal Donor Strain *Sc*_HO19 to facilitate mating (Figure 6A). The pin replica plating procedure outlined in Figure 6B was then followed, which was based on the procedure of Reid *et al.* (2011) but which was modified to also include a step for enrichment of successfully mated strains via growth on a double antibiotic plate selecting for the

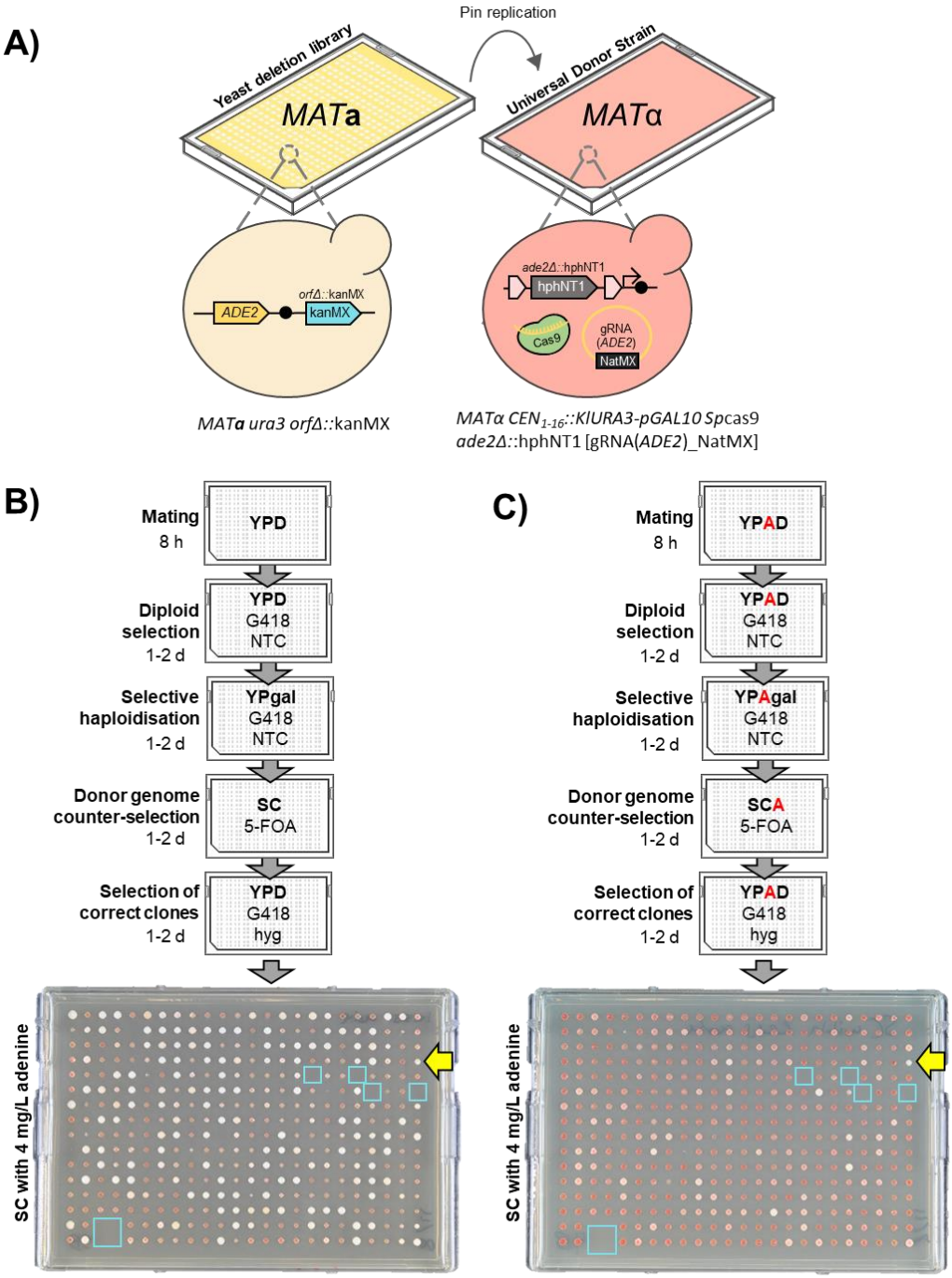


Figure 6. Initial test of a CRI-SPA plating protocol for deletion of the *ADE2* gene in a subset of the yeast deletion collection. (A) Experimental set-up. A subset of the yeast deletion library (plate 9_{384}) was arrayed on a rectangular agar plate and were then transferred by replica pinning onto a lawn of the Universal Donor Strain (UDS) *Sc*_HO19 in order to facilitate mating. The *Sc*_HO19 UDS harbours the *ade2*Δ*::hphNT1* deletion and a gRNA plasmid for targeting the *ADE2* locus. Other relevant genotypic information for the deletion collection strains and the UDS is also specified. (B) and (C) Two different pin replication protocols were used in the transfer of the *ade2*Δ::hphNT1 allele, differing in the amount of adenine supplemented to the growth medium. In (B), medium was prepared with normal levels of adenine whereas plates in (C) were supplemented with 40 mg/L adenine. In each case (B) and (C) shows photographs of final strains generated by CRI-SPA when grown on SC with only 4 mg/L adenine. The position of the *ade2Δ ade3Δ* double mutant is indicated by the yellow arrow, and empty control spots by blue squares.

NatMX marker of the gRNA plasmid (harboured by the donor) and the *kanMX* marker (in the genetic background of the recipient). After the final selection step on G418 and hygromycin, cells were transferred to solid SC medium supplemented with only 4 mg/L of adenine in order to enhance the phenotype of *ade2Δ* cells. The plate was photographed and is shown in Figure 6B.

The position of the *ade3* mutant is highlighted in Figure 6B and, as expected, this mutant remained white. However, an unexpected large number of additional mutants on the plate also displayed a white phenotype rather than the red appearance of an *ade2*Δ clone, indicating that the method was prone to artefacts. The accumulated red pigment in *ade2* mutants are toxic to the cell and mutants therefore have a slowed growth compared to the wild-type (18). Hence, in the mix of cells growing on the plates of the CRI-SPA plating procedure, non-targeted, wild-type cells could quickly outgrow correctly edited *ade2* mutants. Furthermore, although the rate is low, there is a selective pressure for *ade2* mutants to revert their phenotype by mutations elsewhere in the genome that causes the adenine synthesis pathway to stall upstream of Ade2, thereby preventing production of the toxic red pigment. We hypothesised that the large number of white clones observed in Figure 6B were likely wild-type cells, which had outcompeted correctly edited *ade2* cells. To improve the screen, the experiment was repeated and in each selective step, the growth medium was supplemented with extra adenine (40 mg/L) in order to lessen the competitive disadvantage of *ade2* cells. The procedure is outlined in Figure 6C. After completion of the selection protocol, strains were transferred to solid SC medium with reduced amounts of adenine (4 mg/L adenine) to enhance the *ade2* phenotype, see Figure 6C.

Encouraging, using this alternative plating protocol, a white phenotype was again observed for the *ade2 ade3* mutant, whereas the majority of remaining strains on the plate displayed a red or pink phenotype. It was clear that addition of extra adenine had a strong effect and that competition between *ade2* and *ADE2* cells had significant impact on the efficiency of the screen. Additional white clones to the *ade2 ade3* mutant were also present on the plate derived from this alternative plating scheme. As it was uncertain if these mutants represented cases of true genetic interactions with *ADE2* or artefacts of the CRI-SPA procedure, their genetic identity was analysed in closer detail along with additional clones generated in the experiment.

Phenotypical analysis of CRI-SPA generated clones

Cells growing in different positions on the final HTS plate of the replica pinning scheme outlined in Figure 6C, appeared either white, pink or red. Since each patch represents a population of individual recombinants rather than a clone of a single event, the complete patch of cells from two white (*asc1*Δ and *scp160*Δ), two pink (*hse1*Δ and *bud4Δ*) and two red patches (*yer097w*Δ and *atp20*Δ) was individually picked, re-suspended in sterile water, diluted and then added to SC agar plates (4 mg/L adenine) in order to generate single colonies (Supplementary Figure 2-4). The numbers of red and white colonies were recorded (Table 1). Cells originating from red patches on the HTS plate yielded close to 100 % red colonies when grown as single colonies – only cells originating from the O18 patch (*atp20*Δ) did yield a few small, white colonies. Cells originating from pink patches did, unsurprisingly, produce a mix of red and white single colonies when spread out on the SC plates, explaining their pink appearance on the HTS plate. Cells originating from white patches showed a greater fraction of white colonies, but did in fact also yield red colonies when spread onto SC agar plates, indicating that they represented false positives. The mutations in question, *scp160Δ* and *asc*1Δ, are therefore not likely

Table 1 Analysis of single colonies from mixes of cells growing on the final high-throughput screening (HTS) plate of Figure 6C.

Position	Mutation	Appearance on HTS plate	Number of red colonies	Number of white colonies	% red colonies
K20	*asc1Δ*	White	3	67	4.3 %
N15	*scp160Δ*	White	44	77	36%
F2	*hse1Δ*	Pink	177	42	81%
L2	*bud4Δ*	Pink	127	38	77%
I1	*yer097wΔ*	Red	252	0	100%
O18	*atp20Δ*	Red	133	2	99%

to interfere with adenine synthesis, but rather with cellular processes influencing the functionality and/or efficiency of the CRI-SPA method. Indeed, *SCP160* encodes a protein involved in the pheromone mating response (19) and *asc1*Δ mutants have been found to be highly sensitive to 5-fluorouracil (5-FU) (20) - the toxic compound generated when 5-FOA is metabolised in *URA3* cells (21).

Single colonies were further analysed by replica plating onto different selective media. Colonies originating from red patches on the HTS plate consistently displayed the expected phenotype of a correctly manipulated clone - growing both on hygromycin and G418-supplemented plates but not on SC plates lacking either adenine or uracil. This thus indicated the complete loss of donor chromosomes and the replacement of the *ADE2* gene with the disrupted *ade2*Δ::hphNT1 allele in the recipient genetic background. This also led us to believe that the small, white colonies originating from the O18 position (*atp20*Δ) were actually most likely correct in terms of a disrupted *ADE2* locus, but remained white due to petiteness (22). For colonies originating from pink patches, complete loss of UDS chromosomes could also be confirmed as there was no growth on medium lacking uracil, and all colonies further exhibited resistance to both hygromycin and G418 indicating disruption of the *ADE2* locus in the recipient by insertion of the *hphNT1* marker. The occurrence of white colonies was thus puzzling. These white colonies were further able to grow without adenine suggesting a functional synthesis pathway, and when investigated by colony PCR, the presence of both an intact and disrupted *ADE2* locus could be observed (data not shown). Identical observations were made for white colonies originating from white patches on the HTS plate, but these colonies did occasionally also display growth on SC-ura, indicating incomplete loss of donor chromosomes. Together these analyses indicate that patches contained (at least) two different recombinants: red cells that were generated by simple gene conversion, and undesired white cells that appeared to be heterozygous for the *ADE2* locus. Moreover, formation of white cells appeared to be less frequent as compared to formation of red cells, but since white cells have a growth advantage over red ones, they may take over the population during the growth phases of the selection steps.

In a 2008 study by Reid *et al.* leading up to the development of the final SPA method in their later study of 2011, Universal Donor Strains were constructed in which only one of the 16 chromosomes harboured a conditional centromere (23). When mating such a donor strain to a recipient strain and then destabilising the centromere of the modified UDS chromosome by growing on galactose, the study found that in most cases, $2n-1$ cells were not generated as would be

expected, but rather 2*n* diploid cells. These cells were explained to be the result of endoreduplication, a process in which haploinsufficiency of 2*n−1* cells triggers the duplication of the non-paired chromosome in order to restore a diploid genome. This is a known phenomenon, which is also used in the creation of the synthetic yeast genome (24). We hypothesised that endoreduplication might be the cause of the white phenotypes observed in our experiments. Specifically, we envisioned that if endoreduplication of the receptor chromosome takes place in cells prior to the action of Cas9, then one of these chromosomes could escape cleavage and subsequent gene conversion. This would produce heterozygous *ADE2/ade2*Δ*::hphNT1* cells that would survive the CRI-SPA selection procedure. To investigate this matter further, the genomes of a small number of clones from the analysis above were sequenced and their chromosome ploidy level was assessed.

Whole genome sequencing and ploidy measurements of CRI-SPA generated clones

Observations from the phenotypic screen of CRI-SPA generated clones warranted more detailed analysis of their genomic set-ups. Colonies of the following strain backgrounds were included in the analysis: *scp160*Δ, *hse1*Δ and *atp20*Δ and for each strain, both a white and red colony was investigated. Whole-genome DNA sequencing revealed that *K. lactis URA3* marker sequences were absent in the genome of all the sequenced clones, thus confirming complete ablation of donor genome. The *hphNT1* marker sequence was further detected in all the sequenced genomes and for the white clone of *hse1*Δ and both clones of *scp160*Δ, integration of the marker at the *ADE2* locus could also be confirmed, whereas inadequate contig read lengths for other the clones prevented any clear conclusions on the location of the marker. In white clones, a wild-type *ADE2* sequence was also observed, except in the case of the white clone of ***atp20Δ***, which, as suspected, did not contain any mitochondrial DNA and thus displayed a white phenotype due to petiteness. This clone also showed chromosomal rearrangements of Chr III – a phenomenon also observed by others (25) and which was considered unrelated to the Cas9 targeting of the *ADE2* locus, which is located on Chr XV. Moreover, for the red clone of *scp160*Δ, variations in coverage between chromosomes revealed that all chromosomes were present in two copies except Chr V, XII, XII and XVI, which were present in three copies. For the remaining clones, such conclusions regarding ploidy level could not be made.

To further assesses the ploidy level of the sequenced clones, relative amounts of DNA were next determined by staining cells with propidium iodide (PI) and measuring fluorescence by flow cytometry (Figure 7). Comparison to a haploid and diploid reference strain suggested that ***hse1Δ*** and ***atp20Δ*** clones all had a diploid genome. Moreover, the aneuploid genome seen in the sequencing of the red ***scp160Δ*** clone was also observable in this experiment as the clone displayed levels of fluorescence neither matching that of the haploid or diploid reference strain. Similar observations could also be made for the white clone of the ***scp160Δ*** mutant. Upon further inspection of the *SCP160* gene function, it was found that in addition to its role in mating, the *SCP160* gene product is also involved in chromosome segregation and ploidy maintenance, and disruption of the gene causes increased DNA content in the mutant (26). It is possible that is a major reason that the ***scp160Δ*** mutant appeared white on the final HTS plate in our ***ade2Δ*** experiment, as aneuploid clones harbouring additional copies of chromosome XV makes the complete Cas9-targeting of all *ADE2* sequences more difficult.

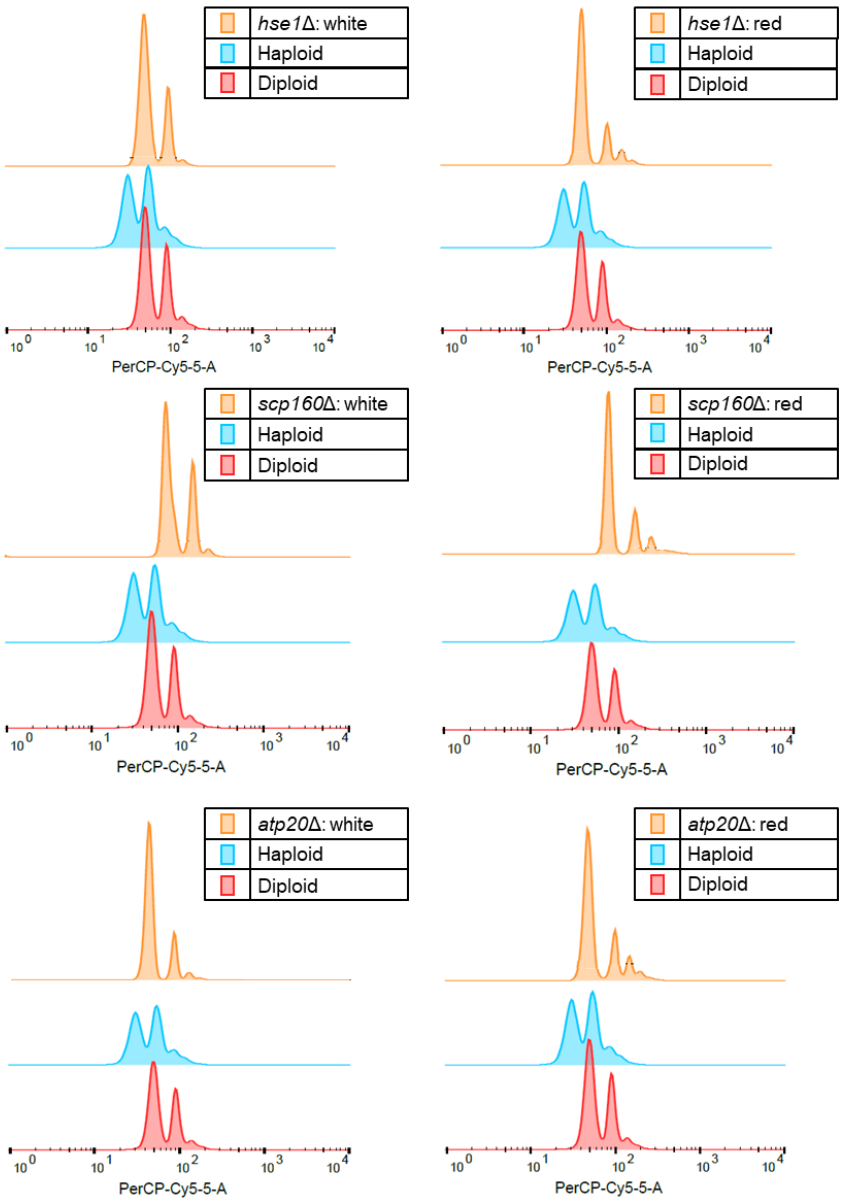


Figure 7. Assessment of ploidy level of CRI-SPA generated strains. Strains originate from the plating procedure in Figure 6C and in each case, a white and a red single colony of the respective strain was analysed. Cells were stained with propidium iodide and the histograms show the distribution of cells in accordance to their relative DNA content when analysed by flow cytometry. Ploidy was assessed by comparing histrograms of each clone to that of a known haploid and diploid reference strain.

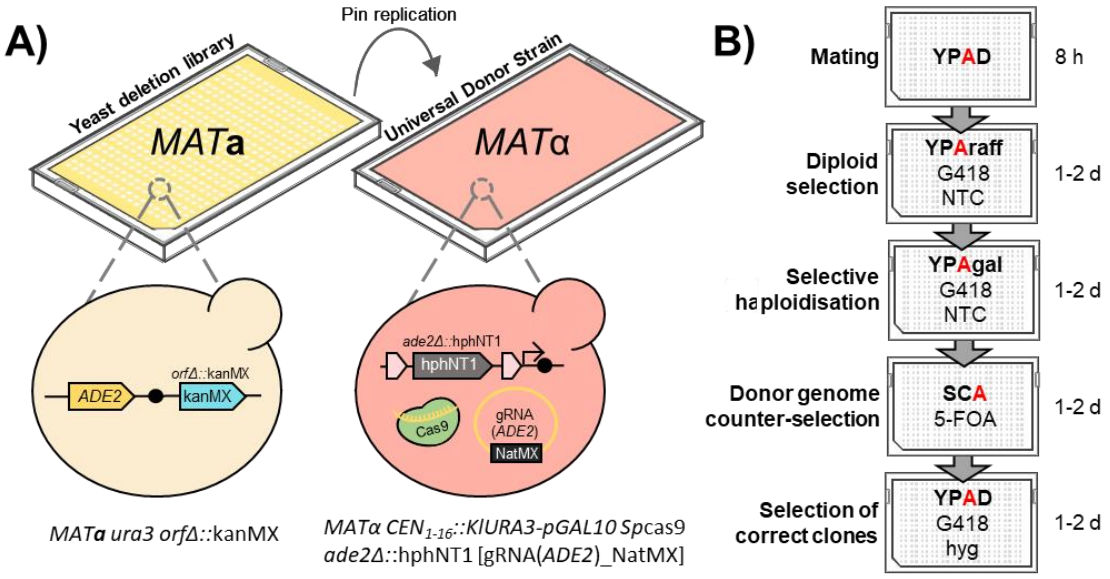
Observations from genome sequencing and ploidy level assessment led to the conclusion that the investigated, CRI-SPA generated clones had a non-haploid genome but which only harboured chromosomes originating from the background of the recipient strain. The *hse1Δ* and *atp20Δ* clones were diploid whereas the *scp160Δ* clones were additionally aneuploid for a number of chromosomes, which can likely be derived to disrupted chromosome segregation in the mutant. Our observations thereby verified the initial hypothesis of endoreduplication. This process causes complete duplication of the recipient genome and white *ADE2*/ade2Δ::*hphNT1* clones are thereby a result of inefficient Cas9

targeting of both chromosomes. Although such cells are likely initially in minority, they have a growth advantage over correctly edited *ade2* cells and may cause issues if taking over the heterogeneous population of cells growing on the plate. As noted by Reid *et al.* (2008), duplication of the recipient chromosomes, still enables SPA-based screens to link phenotypes to genotypes of the deletion library, as the genetic set-up of the recipient remains the same although in a diploid cell. In the case of CRI-SA, it does however, require an efficient Cas9 to target both chromosomes of the recipient in order to avoid cells that are heterozygous for the transferred allele.

Establishing an improved CRI-SPA plating protocol for genetic interaction studies

After observing duplication of chromosomes and heterogeneous Cas9 targeting of these in recipient strains, changes to the selective plating procedure were introduced. In the original set-up, mated strains were grown on medium supplemented with glucose before being transferred onto galactose-containing plates for inducing the loss of donor chromosomes. Glucose is the preferred substrate for yeast and growth on this carbon source represses the expression of genes needed for utilisation of less favoured carbon sources such as galactose. We hypothesised that inefficient induction on galactose due to glucose repression may be problematic, causing unsynchronised loss of UDS chromosomes and endoreduplication. In an attempt to avoid this, a new plating procedure was tested in which diploid selection after mating was carried out using growth medium in which raffinose was the sole sugar, thus metabolising and removing any remaining glucose before plating onto galactose. A small test experiment using a manual pin replication procedure was carried out with a smaller subset of the yeast deletion library, to which the *ade2*Δ::*hphNT1* allele was again transferred. This protocol gave promising results as it yielded a significant fraction of red clones even in the case of not supplying additional adenine throughout the procedure (data not shown). However, a few strains were picked from the final CRI-SPA plate and streaked out into single colonies and their DNA contents were analysed by PI staining and flow cytometry. For most clones, a diploid genome could still be observed (Supplementary Figure 5). Thus, it appeared that endoreduplication had again caused duplication of the recipient genome. Nevertheless, the positive effect that the growth step on raffinose seemed to have on the ability of Cas9 to correctly edit both chromosomes of the recipient genome was encouraging.

The experiment was repeated in higher throughput using the same automated pinning procedure as previously, transferring the *ade2Δ::hphNT1* deletion to the same subset of the yeast deletion collection as in Figure 6. This time, however, a plate set-up was used in which the library strains were arrayed in a 32 by 48 grid (1536 density format) and in which each strain was present in quadruplicate in a 2 by 2 grid. Biological replicates is key to any HT screen and enables increased accuracy of analysis, as false positives more easily can be identified and disregarded when comparing replicates. As outlined in Figure 8A, the arrayed strains were mated to donor strain *Sc*_HO19 and was then taken through the plating procedure presented in Figure 8B. As previously, all plates were supplemented with extra adenine. After growth on the final selective plate, cells were transferred to synthetic medium supplemented with lower concentrations of adenine to enhance the phenotype of *ade2* mutants. The plate was photographed and is displayed in Figure 8C.



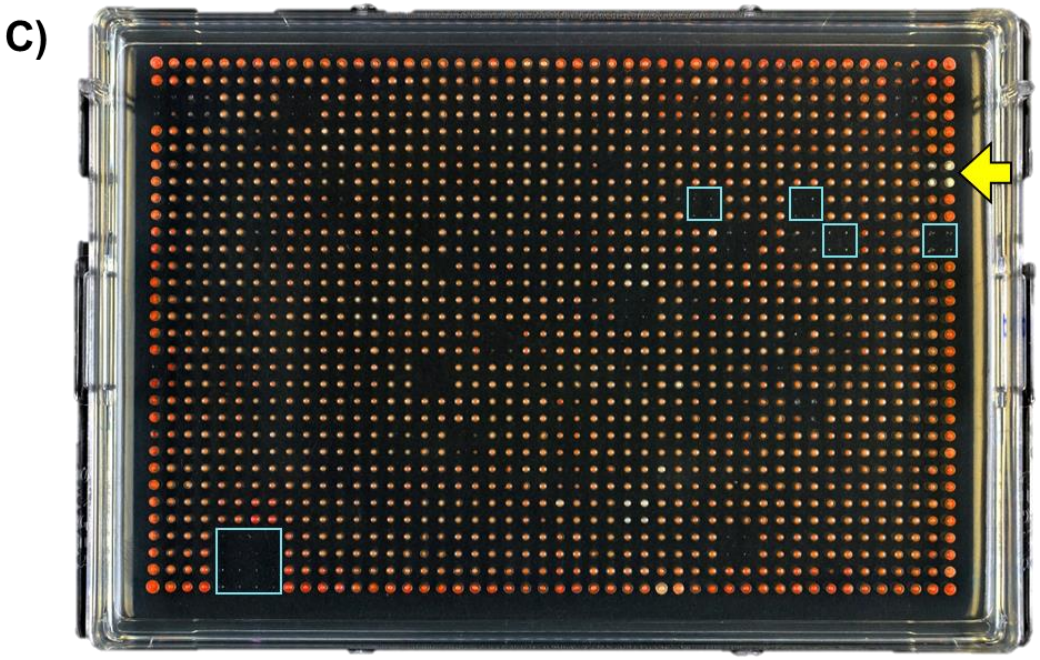


Figure 8. High-efficiency pin replica plating procedure for the high-throughput deletion of *ADE2* in a subset of the yeast deletion collection. (A) Experimental set-up. A subset of the yeast deletion library (plate 9_{384}) was arrayed on a rectangular agar plate and was then transferred by pin replication onto a lawn of the Universal Donor Strain (UDS) *Sc*_HO19 to facilitate mating. The *Sc*_HO19 UDS harbours the *ade2*Δ::hphNT1 deletion and a gRNA plasmid for targeting the *ADE2* locus. Other relevant genotypic information of the deletion library and the UDS is also specified. (B) Pin replication protocol. Additional adenine (40 mg/L) is supplemented throughout the replication procedure in order to lessen the growth disadvantage of *ade2* cells. After mating, cells are transferred to a diploid selection plate in which raffinose is the sole sugar in order to avoid glucose repression in the following induction step on galactose. (C) Final strains grown on SC with only 4 mg/L adenine in order to enhance the red phenotype of *ade2* cells. Each strain is present in quadruplicate in a 2 by 2 array. The yellow arrow indicates the position of the *ade2Δ ade3Δ* double mutant and blue squares the positions of empty control spots.

As seen in Figure 8C, efficiency of transfer was significantly higher than in previous experiments and the red phenotype could be observed for the majority of strains growing on the plate. Moreover, running the experiment with biological replicates enabled false positives to be more easily disregarded when observing different phenotypes between replicates. Only two strains displayed a white phenotype for all four biological replicates. One of these strains (position N15) carried a mutation in the *SCP160* gene. This mutation also yielded white cells in the previous experiments and as outlined above, it presented a case of a false positive as red phenotypes could be observed when the mutant was streaked out into single colonies. The second strain was again the double *ade2Δ ade3Δ* mutant, which, as expected, did not produce the red pigment and which again demonstrated the accuracy of the CRI-SPA method in identifying this genetic interaction. A very pale pink colour was also seen for the mutant carrying a mutation of the *MET6* gene (position G15). *MET6* encodes an enzyme part of the methionine synthesis pathway, catalysing the conversion of homocysteine to methionine. Interestingly, there is a link between synthesis of adenine and methionine in this reaction step (27). In the reaction, the co-factor 5-methyl-THF donates a one-carbon group to homocysteine and is in the process converted to THF. The co-factor then cycles back to its methylated forms, 5-methyl-THF and 5,10-methylene-THF. The latter, 5,10-methylene-THF, is required in the synthesis of adenine where it is converted to 10-formyl-THF by Ade3 (Figure 4). There is hence a connection between the synthesis of adenine and the synthesis of methionine in the cell via THF metabolism (Figure 9) and reduced levels of 5,10 methylene-THF may explain the limited production of red pigment in the *ade2Δ met6Δ* double mutant. This double mutant was not observed in earlier screens (Figure 6B and C) and it possible that inefficient loss of donor chromosomes in earlier set-ups masked its phenotype.

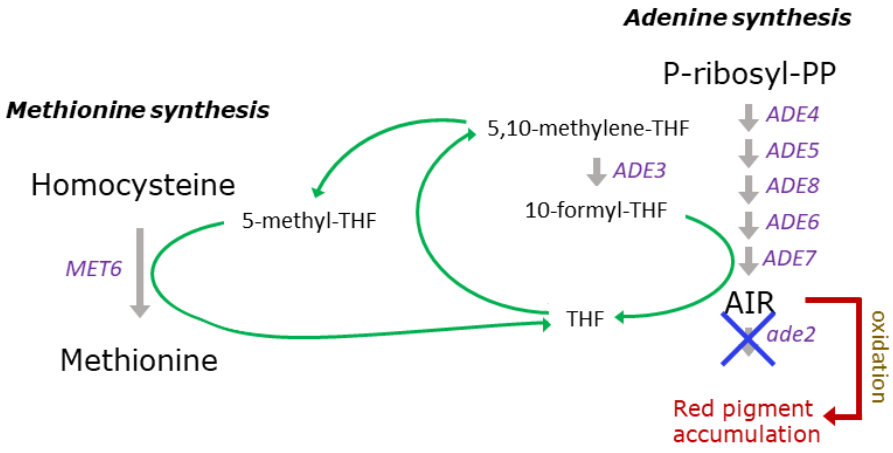


Figure 9. Simplified outline of the cross-talk between the methionine and adenine synthesis pathway via tetrahydrofolate (THF) metabolism

In addition to observing the colour of mutants on the final HTS plate, several mutant strains of which all four biological replicates failed to grow on the plate in Figure 8C could also be observed, potentially indicating a synthetic lethal interaction in the double mutants. Three of these mutants also yielded empty or completely white positions on the final plates of Figure 6B and 6C: *vma13*Δ, *thr1*Δ, and *mtc1*Δ. These mutants were, however, also found to be non-growing in later CRI-SPA based screens, which did not involve the transfer of the *ade2*Δ deletion, and were thus disregarded as

false positives. Looking into each their respective gene function, no obvious effects on mating efficiency were found. The *vma13*Δ mutant, does however, in similarity to the *asc1Δ* mutant discussed earlier, display increased sensitivity to 5-FU (20). It may be possible that although *URA3*-tagged chromosomes of the UDS have been lost, there is still sufficient *URA3*-encoded enzyme present in these mutants to cause inhibition when grown on 5-FOA.

Discussion

The yeast deletion collection is a valuable resource available to the yeast research community. However, genetic manipulation of the ca. 4800 haploid deletion mutants remains a challenge and requires methods that can be carried out in high throughput. Using a mating-based, CRISPR-Cas9 assisted strategy to transfer a genetic element from a donor strain into the mutants of the deletion collection makes CRI-SPA a suitable method for this task as it can easily be automated and carried out in high-throughput using a robotic pin replication procedure.

CRI-SPA is, in contrast to other mating-based methods for manipulation of the deletion collection, independent of complex meiotic processes and sporulation procedures to haploidise mated cells. The genetic background of the deletion collection, S288C, has a poor sporulation capacity and a sporulation-free method such as CRI-SPA may therefore be advantageous in terms of efficiency and throughput of the method, and it significantly shortens the plating protocol, here requiring less than a week to complete. Furthermore, meiotic recombination and scrambling of donor and recipient genome is avoided with CRI-SPA and the method can therefore be used with genetically diverse, non-isogenic strain libraries in which one wishes to preserve the genetic background of the recipient strain. This is a feature that we address further and that we take advantage of in Chapter 4 of this thesis. A non-meiotic procedure also means that CRI-SPA may be used for interspecies crosses between a UDS and haploid derivatives of other *Saccharomyces* species, which are able to mate to *S. cerevisiae* but which typically produce sterile hybrids unable to go through meiosis, examples including *S. paradoxus*, *S. kudriazevii* and *S. bayanus* (28). Finally, independency of meiosis means that a CRI-SPA system could ultimately be extended to microbial species lacking a sexual cycle, such as for example the industrially important filamentous fungi *Aspergillus niger* for which a sexual cycle remains elusive (29). In the laboratory, two genetically different strains of *A. niger* may, however, be fused (anastomosis) to form diploids, and then haploidised in a process called the parasexual cycle (30), hence setting the stage for CRI-SPA.

CRI-SPA is also benefitted by its incorporation of CRISPR-Cas9 technology, which opens up possibilities for multiplexing. We have demonstrated that CRI-SPA can be used to conduct genetic interaction screens by the transfer of a single gene deletion to the deletion collection, yielding double mutants. However, using the proper UDS and multiple gRNAs, higher order interactions could easily be studied by the simultaneous transfer of several deletions and could have useful applications in fundamental research uncovering the connectivity and interaction between networks of genes. Similar studies are also possible with the SGA method (31), but will get increasingly difficult with increasing number of transferred mutations as it requires co-segregation of mutated chromosomes during meiosis.

In addition, the use of CRISPR-Cas9 may allow marker-free integration of DNA by CRI-SPA, which could be an advantage in many experiments, for example in the transfer of point mutations. However, in this study, we chose to transfer our mutant allele along with a dominant marker, as it safeguards against some of the false positives one would

expect from a high-throughput set-up in which heterogeneous mixes of cells are transferred between plates. We therefore note that multiplexing with CRI-SPA may be limited by the number of available markers one can use for the transfer of each individual genetic element. However, in the case of using CRI-SPA together with the yeast deletion collection, several auxotrophic markers are available for use as the strains of the collection are *his*$^-$, *leu*$^-$ and *met*$^-$ (in addition to *ura*$^-$). Moreover, exciting split-enzyme-marker systems have recently been demonstrated, in which antibiotic resistance marker genes are portioned into separate sequences encoding different parts of the enzyme conferring resistance to the antibiotic. Only when all the separate split-marker sequences are present do the individual peptides fuse into a functional enzyme, allowing survival when grown in media supplemented with the antibiotic (32). A system like this would be highly interesting to combine with CRI-SPA for efficient and multiplex editing.

In our studies, we observed duplicated genomes of the recipient strains after being modified by the CRI-SPA method. We find it likely that this is a result of endoreduplication and it may cause issues if generated clones are heterozygous for the allele (or other genetic feature) that is being transferred. Although efficient, CRISPR-Cas9 targeting is rarely 100 % effective. In most gene editing scenarios, this is not a problem as the majority of generated clones will be correct and easily identified by screening a small number of clones. However, in a high-throughput replica plating procedure such as the one employed for CRI-SPA, cells that have escaped Cas9 targeting and which harbour one modified and one non-modified chromosome may, although initially present in minority, gradually take over the heterogeneous population of cells growing on the plates, given that they have a growth advantage over correctly edited cells. This issue was clearly evident in our initial *ade2* experiments, where correctly edited *ade*$^-$ cells had a growth disadvantage and were therefore frequently outcompeted by *ADE*$^+$ cells. In this case, there was a simple solution at hand, as supplying elevated concentrations of adenine significantly lessened the competitive disadvantage of *ade*$^-$ cells. However, this type of strategy to relieve growth defects may not always be applicable and one may have to think of alternative solutions if the transfer of a gene deletion causes severe impairments to fitness. Temporary expression of the targeted gene from the same plasmid which harbours the gRNA may be a solution, but measures for ensuring plasmid loss at the end of the transfer protocol would then be needed. In either case, Cas9 efficiency is key to the efficiency of CRI-SPA based transfers, and we found that growing strains on raffinose before replicating onto galactose greatly increased the number of correct clones in our *ade2* experiment. Although resulting clones were diploid, this procedure seemed to allow more efficient Cas9 targeting of both chromosomes. The reason for this is yet not obvious to us. In combination with supplying elevated levels of adenine, the replica plating scheme with the added step on raffinose enabled highly efficient transfer of the *ade2*Δ::*hphNT1* allele to the recipient strains.

CRI-SPA consistently identified the known genetic interaction of *ADE2* and *ADE3* in our screens. Furthermore, the screen could identify the link between adenine and methionine synthesis by observation of suppressing genetic interactions in the *ade2Δ met6Δ* double mutant. Additionally, a few false positives were also encountered in the case of the *scp160*Δ mutant, which appeared white on the HTS plate but produced red single colonies, and the *vma13Δ*, *mtc1Δ* and *thr1Δ* mutants, which failed to grow on the final HTS plate but which phenotypes were not caused by a synthetic lethal interaction with *ade2Δ*. We find it likely that these mutations have a general effect on the functionality of the CRI-SPA procedure and for some of them, connections to impaired mating , ploidy maintenance and increased sensitivity to

5-FU were found. The use of high-throughput methods, especially together with the yeast deletion collection, which has its own pitfalls in terms of wrongly annotated ORFs, second-site genomic mutations in addition to the disrupted ORFs (33,34), aneuploidies (35), will inescapably generate a number of false positives and negatives. Caution is therefore always prompted before drawing conclusions from a HT screen and an identified hit should be verified by introduction of the mutation in a background strain in which the phenotype then can be confirmed.

The CRI-SPA method is adapted for use with haploid recipient strains able to mate to the UDS and which are further *ura*⁻. We address these limitations in Chapter 4 of this thesis, where we demonstrate the use of CRI-SPA together with non-isogenic, diploid strains which have been made compatible with the method. In its current set-up, CRI-SPA is, however, not only compatible with the haploid versions of the deletion library but also with other useful libraries including the GST-tagged ORF overexpression library (36), and the DAmP collection for studies of essential genes (37). In addition to genetic interaction screens applied in fundamental research, we see great potential in using CRI-SPA in yeast cell factory research. By transferring a biochemical production pathway to any of the genome-wide mutation libraries listed above, genetic designs that return increased or decreased production of the compound in question could be identified and the information can be implemented in a metabolic engineering strategy. This application of CRI-SPA in cell factory optimisation is investigated in Chapter 3 of this thesis.

Methods

Strains and media

All strains used and constructed in this work are listed in Table 2. The universal donor strains W8164-2B and W8164-2C of the SPA method (11) and which were used to construct new donor strains for CRI-SPA were kindly provided by Rodney Rothstein (Department of Genetics and Development, Columbia University, USA). The yeast genome-wide deletion collection was obtained from Invitrogen. For the experiments outlined in this study, plate Δp9 (384 density format) was used (see positions of mutant strains in Supplementary Figure 1).

Table 2. Strains used and constructed in this study.

Strain name	Genotype	Source
Universal Donor Strains		
W8164-2B	*MAT*α CEN$^{1\text{-}16}$::p*GAL1-KlURA3 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5*	(11)
W8164-2C	*MAT***a** CEN$^{1\text{-}16}$::*pGAL1-KlURA3 can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5*	(11)
*Sc*_HO10	*MAT*α CEN$^{1\text{-}16}$::*pGAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1-Sp*Cas9-t*CYC1*-loxP-*KlLEU2-loxP*	This study
*Sc*_HO11	*MAT***a** CEN$^{1\text{-}16}$::p*GAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1-Sp*Cas9-t*CYC1*-loxP-*KlLEU2*-loxP	This study
*Sc*_HO19	*MAT*α CEN$^{1\text{-}16}$::p*GAL1-KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1-Sp*Cas9-t*CYC1*-loxP-*KlLEU2-loxP KlURA3(dw XII-5 IS)*	This study
Strains of the deletion library	*MAT***a** *orf*Δ::kanMX4 *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 TRP1 CAN1 LYS2 ADE2*	(3,4)

Strains were preserved in YPD glycerol stocks (20% (v/v)) at - 80°C. Throughout the study, strains were grown in YP (10 g/L yeast extract, 20 g/L peptone) using either glucose (20 g/L), galactose (20 g/L) or raffinose (20 g/L) as carbon source or in synthetic complete (SC) medium prepared as described by (38) but with 60 mg/L L-leucine. Galactose and raffinose solutions were sterilised by filtration. For growth on solid medium, 20 g/L agar was added. For selection, plates were supplemented with G418 (200 mg/L), nourseothricin (NTC) (100 mg/L) and hygromycin (hyg) (200 mg/L) in different combinations. When grown in liquid, only half of the concentration of the respective antibiotic was used. For selection on SC, the appropriate amino acid was dropped out. To enhance the phenotype of *ade2* cells, SC medium with only 4 mg/L adenine was prepared. To prevent *ade2* cells from reverting, growth media were supplemented with 40 mg/L adenine. For *URA3* counter selection, SC plates were supplemented with 5-fluoroorotic acid (5-FOA) (1 g/L) and uracil (30 mg/L). For cloning purposes, *Escherichia coli* DH5α was grown in LB medium (39) supplemented with 100 mg/L ampicillin and in the case for growth on solid medium, 20 g/L agar was added.

Design of gRNA sequences

gRNA target sequences in this study were generated using the Benchling CRISPR tool (https://benchling.com). Off-target effects were minimized by using the S288C reference genome and the algorithm from (40).

General cloning procedures

Plasmids used and constructed in this study are presented in Table 3. The pMEL12 plasmid (41) was obtained from EUROSCARF (Frankfurt, Germany). Primers are listed in supplementary Table 1 and were ordered from Integrated DNA Technologies (Leuven, Belgium). All cloning was done using uracil-specific excision reagent (USER™) enzyme from New England Biolabs and was carried out as previously described (42,43). DNA polymerases were purchased from Thermo Scientific and used according to the supplier's instructions. Standard PCR amplifications were run using Phusion Hot Start II High-Fidelity DNA Polymerase. Amplification of DNA for USER cloning was carried out using Phusion U Hot Start DNA Polymerase. In the case of colony PCR, Taq DNA Polymerase was used. Purification of DNA from PCR reactions or from agarose gel bands was done using the illustra GFX PCR DNA and Gel Band Purification Kit (GE Lifesciences). All plasmids were verified by sequencing at Eurofins MWG Operon (Germany).

Table 3. Plasmids used and constructed in this study.

Plasmid name	Genetic elements	Yeast marker	Type	Source
pCfB257	-	*KlLEU2*	Integrative, X-3	(44)
pCfB390	-	*KlURA3*	Integrative, XI-3	(44)
pCfB2311	p*SNR52*-gRNA(*ADE2*)-t*SUP4*	NatMX	2µ	(45)
pCfB2312	p*TEF1*-SpCas9-t*CYC*	kanMX	CEN/ARS	(45)
pCfB3050	p*SNR52*-gRNA(XII-5)-t*SUP4*	NatMX	2µ	(46)
Ant_E106	p*SNR52*-gRNA(no target)-t*SUP4*	NatMX	2µ	(47)
pMEL12	p*SNR52*-gRNA(*CAN1*)-t*SUP4*	hphNT1	2µ	(41)
pHO8	p*TEF1-Sp*Cas9-t*CYC1*	*KlLEU2*	Integrative, X-3	This study
pHO22	p*SNR52*-gRNA(dw *ADE2*)-t*SUP4*	NatMX	2µ	This study
pHO24	p*SNR52*-gRNA(*ADE2*)-t*SUP4*	NatMX	2µ	This study

Plasmid construction

To construct the plasmid for insertion of Cas9 at the chromosomal integration site X-3 (48), the Cas9 gene was amplified from plasmid pCfB2312 (45) with primers HOP89 and HOP90. The fragment was then cloned into the integrative plasmid pCfB257 (44), which prior to the cloning had been linearized by primers HOP91 and HOP92, thereby yielding plasmid pHO8.

For construction of a gRNA plasmid targeting an intergenic region downstream the ADE2 locus, the target sequence in pCfB3050 (44) was excluded by linearising the plasmid with uracil-containing primers HOP60 and HOP61. The linearised vector was then treated with DpnI and gel purified. A cloning cassette containing the new 20 bp targeting sequence was generated by annealing the two single stranded oligos HOP173 and HOP174 to form a double-stranded fragment with a 5' and 3' single-stranded end, each matching the overhangs of the linearised and USER-treated gRNA plasmid backbone. The annealing was done by mixing equal volumes of each oligo (100 µM), boiling for 5 minutes and then allowing the solution to cool to room temperature. The target sequence cassette was USER cloned into the linearised pCfB3050 backbone to yield new gRNA plasmid pHO22. gRNA plasmid pHO24 targeting *ADE2* was constructed in a similar manner but using a double-stranded fragment generated from annealing HOP190 and HOP191.

Strain construction

Strains used and constructed in this study are listed and described in Table 2 above. All yeast transformations were carried out using the LiAc / SS Carrier DNA / PEG method (49). When antibiotic resistance genes were used as markers, cells were recovered for 2 hours in YPD before being plated onto selective media. For all transformations, negative controls were included, where no DNA was added to the transformation mix. In each case between 200 -500 ng plasmid DNA was used and between 750 – 1500 ng of linear repair substrates. All linear DNA generated by PCR was gel purified prior to transformation. All strains was verified by colony PCR to ensure correct integration at the intended chromosomal locus.

To construct new donor strains carrying the Cas9 gene, the pHO8 plasmid was linearised by digestion with FastDigest *NotI* (Thermo Scientific) and gel purified. The linearized plasmid was then used to transform W8164-2B and W8164-2C. Transformants were selected on SC-leu plates yielding the new donor strains *Sc*_HO10 (*MAT*α) and *Sc*_HO11 (*MAT***a**), respectively.

For construction of the universal donor strain *Sc*_HO19 (*ade2Δ*::hphNT1), a repair substrate for disruption of *ADE2* by insertion of the *hphNT1* (hygR) cassette, was generated by PCR using plasmid pMEL12 and primers HOP64 and HOP65 each with a 45 bp non-binding tail corresponding to the flanking sequence up- respectively downstream of the *ADE2* ORF. The fragment was gel purified was then used to transform strain *Sc*_HO10 along with gRNA plasmid pCfB2311. Transformants were selected on YPD supplemented with NTC and hygromycin. After curing this intermediate strain from the pCfB2311 plasmid, the strain was further modified to harbour an additional *KlURA3* marker by transformation with gRNA plasmid pHO22 targeting an intergenic region downstream the *ade2Δ*::hphNT1 locus, and a *KlURA3* repair substrate. For this repair substrate, the *KlURA3* marker was amplified using plasmid pCfB390 (44) and primers HOP175 and HOP176 with 40 bp non-binding tails corresponding to the upstream respectively downstream

flanking sequences of the targeted chromosomal site. The fragment was treated with DpnI and gel purified before co-transformation with pHO22. Transformants were selected on YPD supplemented with NTC, yielding strain *Sc*_HO19.

CRI-SPA high-throughput pin replication protocol

Automatic pin replication was carried out using the high-throughput pinning robot ROTOR HDA from Singer Instruments (United Kingdom), along with replica pinning pads (RePads) and rectangular petri dishes (PlusPlates) from the same company.

For parallel mating of the Universal Donor Strain to the strains of the yeast deletion library, a single colony of the UDS was inoculated in 5 mL YPD supplemented with NTC in order to maintain selective pressure for the gRNA plasmid. The UDS was grown overnight and 0.7 mL of the culture was then spread evenly on top of a rectangular YPD agar plate using sterile glass beads in order to generate a lawn of UDS cells. Once the lawn had been allowed to briefly dry in, the arrayed strains of the deletion collection was pinned on top and mating was allowed to proceed for 8 hours. Strains were then pin replicated onto a series of selective plates as outlined in Figure 6 and 8, each time allowing at least 24 hours of growth on each plate.

Whole genome sequencing

Genomic DNA of CRI-SPA generated clones was isolated using the Qiagen 100/G Kit (Qiagen, Hilden, Germany) according to the supplier's instructions. Isolated DNA was quantified using a Qubit® Fluorometer 2.0 with Qubit dsDNA BR Assay kit. The TruSeq DNA PCR-free library preparation workflow was used to obtain 300 cycle paired-end libraries with an average insert size of 550 bp for all samples except for the white clone of *atp20*Δ. Due to lower DNA concentrations, a protocol for generating a library of an average size of 350 bp was followed for this sample. Whole genome sequencing was then carried out using the MiSeq Reagent Kit v3 and an in-house MiSeq sequencer (Illumina, San Diego, CA).

All samples were de novo assembled by applying SPAdes assembler (version 3.9.0) (50). S288C ORFs and integrated ORFs (*hphNT1* and *KlURA3*) were aligned by nucmer from the MUMmer package (version 3.1) (51) to the assembled contigs to identify the presence and location of integration cassettes. The Magnolya algorithm (52) was used to detect chromosomal copy number variation in every sample. For analysis of single nucleotide polymorphisms, sequence reads were mapped to the S288C genome by using the Burrows-Wheeler alignment tool (53) and processed with Pilon (54). Identified single-nucleotide variations were inspected with the Integrated Genomics Viewer (IGV) (55).

Propidium iodide DNA staining and analysis of ploidy level by flow cytometry

Analysis was carried out using the procedure of Stovicek *et al.* (45). Strains were grown in 1 mL YPD in cell cutivation tubes for 2 days with shaking. Cells were then harvested by centrifugation and washed with 1 volume of SSC buffer (150 mM sodium chloride, 15 mM sodium citrate; pH = 7). To fix the cells, the cell pellets were resuspended in 1 volume of a SSC/70 % ethanol solution and samples were then stored at -20°C overnight. The fixed cells were spun down and re-suspended in 1 volume of SSC buffer containing 0.25 mg/mL RNase A and were incubated overnight at 37 °C. Samples were then treated with 1 mg/mL proteinase K for 1 hour at 37° C. Cells were then again spun down and stained by re-

suspending in 1 volume of SSC buffer with 10 µg/mL propidium iodide (PI). Stained cells were kept at 4° C until analysis. Samples very briefly sonicated and a BD LSR Fortessa flow cytometer was then used to analyse the cells, using a blue 488 nm laser (50 mW power) and a 695/40 nm bandpass filter. Histograms were acquired in linear mode and for each sample, fluorescence was recorded from 10,000 cell counts. The FlowJoX software was used to generate the histrogram plots.

References

1. Botstein D, Chervitz S, Cherry M. Yeast as a Model Organism. 1997;277(5330):1259–60.
2. Foury F. Human genetic diseases : a cross-talk between man and yeast. Gene. 1997;195:1–10.
3. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science. 1999;285(5429):901–6. A
4. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. Nature. 2002;418(6896):387–91.
5. Giaever G, Nislow C. The yeast deletion collection: A decade of functional genomics. Genetics. 2014;197(2):451–65.
6. Nislow C, Lee AY, Allen PL, Giaever G, Smith A, Gebbia M, *et al.* Genes required for survival in microgravity revealed by genome-wide yeast deletion collections cultured during spaceflight. Biomed Res Int. 2015;2015:1–10.
7. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science. 2001;294(5550):2364–8.
8. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, *et al.* The Genetic Landscape of a Cell. Science (80-). 2010;327(5964):425–31.
9. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, *et al.* A global genetic interaction network maps a wiring diagram of cellular function. Science (80-). 2016;353(6306).
10. Ben-Ari G, Zenvirth D, Sherman A, David L, Klutstein M, Lavi U, *et al.* Four linked genes participate in controlling sporulation efficiency in budding yeast. PLoS Genet. 2006;2(11):1815–23.
11. Reid RJD, Gonza S, Sunjevaric I, Alvaro D, Ciccone S, Wagner M, *et al.* Selective ploidy ablation , a high-throughput plasmid transfer protocol , identifies new genes affecting topoisomerase I – induced DNA damage. Genome Res. 2011;477–86.
12. Hill A, Bloom K. Genetic manipulation of centromere function. Mol Cell Biol. 1987;7(7):2397–405.
13. Chlebowicz-Śledziewska E, Śledziewski AZ. Construction of multicopy yeast plasmids with regulated centromere function. Gene. 1985;39(1):25–31.
14. Cheeseman IM, Drubin DG, Barnes G. Simple centromere, complex kinetochore: Linking spindle microtubules and centromeric DNA in budding yeast. J Cell Biol. 2002;157(2):199–203.
15. Boeke JD, Trueheart J, Natsoulis G, Fink GR. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. 1987;154(C):164–75.
16. Silver JM, Eaton NR. Functional blocks of the ad1 and ad2 mutants of *Saccharomyces cerevisiae*. Biochem Biophys Res Commun. 1969;34(3):301–5.
17. ROMAN H. Studies of gene mutation in Saccharomyces. Cold Spring Harb Symp Quant Biol. 1956;
18. Ugolini S, Bruschi C V. The red / white colony color assay in the yeast *Saccharomyces cerevisiae* : epistatic

growth advantage of white ade8-18 , ade2 cells over red ade2 cells. 1996;485–92.

19. Guo M, Aston C, Burchett SA, Dyke C, Fields S, Rajarao SJR, *et al.* The yeast G protein α subunit Gpa1 transmits a signal through an RNA binding effector protein Scp160. Mol Cell. 2003;12(2):517–24.

20. Gustavsson M, Ronne H. Evidence that tRNA modifying enzymes are important in vivo targets for 5-fluorouracil in yeast. 2008;666–74.

21. Boeke JD, La Croute F, Fink GR. A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. MGG Mol Gen Genet. 1984;197(2):345–6.

22. Lai-Zhang J, Xiao Y, Mueller DM. Epistatic interactions of deletion mutants in the genes encoding the F1-ATPase in yeast *Saccharomyces cerevisiae*. EMBO J. 1999;18(1):58–64.

23. Reid RJD, Sunjevaric I, Voth WP, Ciccone S, Du W, Olsen AE, *et al.* Chromosome-scale genetic mapping using a set of 16 conditionally stable *Saccharomyces cerevisiae* chromosomes. Genetics. 2008;180(4):1799–808.

24. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, *et al.* Design of a synthetic yeast genome. Science (80-). 2017;355(6329):1040–4.

25. Kok S De, Nijkamp JF, Oud B, Roque FC, Ridder D De, Daran J, *et al.* Laboratory evolution of new lactate transporter genes in a jen1 D mutant of *Saccharomyces cerevisiae* and their identification as ADY2 alleles by whole-genome resequencing and transcriptome analysis. 2012;(2).

26. Wintersberger U, Kühne C, Karwan A. Scp160p, a new yeast protein associated with the nuclear membrane and the endoplasmic reticulum, is necessary for maintenance of exact ploidy. Yeast. 1995;11(10):929–44.

27. Kresnowati MTAP, Van Winden WA, Almering MJH, Ten Pierick A, Ras C, Knijnenburg TA, *et al.* When transcriptome meets metabolome: Fast cellular responses of yeast to sudden relief of glucose limitation. Mol Syst Biol. 2006;2.

28. Greig D, Borts RH, Louis EJ, Travisano M. Epistasis and hybrid sterility in Saccharomyces. 2002;(April):1167–71.

29. Pel HJ, De Winde JH, Archer DB, Dyer PS, Hofmann G, Schaap PJ, *et al.* Genome sequencing and analysis of the versatile cell factory Aspergillus niger CBS 513.88. Nat Biotechnol. 2007;

30. Pontecorvo G. The Parasexual Cycle in Fungi. Annu Rev Microbiol. 1956;

31. Kuzmin E, VanderSluis B, Wang W, Tan G, Deshpande R, Chen Y, *et al.* Systematic analysis of complex genetic interactions. Science (80-). 2018;360(6386).

32. Jillette N, Du M, Zhu JJ, Cardoz P, Cheng AW. Split selectable markers. Nat Commun. 2019;10(1).

33. Grünenfelder B, Winzeler EA. Treasures and traps in genome-wide data sets: Case examples from yeast. Nat Rev Genet. 2002;3(9):653–61.

34. Scherens B, Goffeau A. The uses of genome-wide yeast mutant collections. Genome Biol. 2004;5(7):1–8.

35. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, *et al.* Widespread aneuploidy revealed by DNA microarray expression profiling. 2000;25(july):333–7.

36. Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, *et al.* Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell. 2006;21(3):319–30.

37. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat Methods. 2008;5(8):711–8.

38. Sherman F, Fink G., Hicks J. Laboratory Course Manual for Methods in Yeast Genetics. Cold Spring Harb Lab. 1986;

39. BERTANI G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. J Bacteriol. 1951;62(3):293–300.

40. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34(2):184–91.

41. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NGA, van den Broek M, *et al.* CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res. 2015;15(2):1–15.

42. Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA. USER fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res. 2007;35(7):0–5.

43. Nour-Eldin HH, Hansen BG, Nørholm MHH, Jensen JK, Halkier BA. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res. 2006;34(18).

44. Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, *et al.* EasyClone: Method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. FEMS Yeast Res. 2014;14(2):238–48.

45. Stovicek V, Borodina I, Forster J. CRISPR-Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae* strains. Metab Eng Commun. 2015;2:13–22.

46. Jessop-Fabre MM, Jakočiūnas T, Stovicek V, Dai Z, Jensen MK, Keasling JD, *et al.* EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. Biotechnol J. 2016;11(8):1110–7.

47. Coumou H. Engineering baker's yeast for the production of aromatic plant-derived compounds. Technical University of Denmark; 2016.

48. Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH, *et al.* Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metab Eng. 2012;14(2):104–11.

49. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc / SS carrier DNA / PEG method. Nat Protoc. 2008;2(1):31–5.

50. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;

51. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, *et al.* Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2).

52. Nijkamp JF, Van Den Broek MA, Geertman JMA, Reinders MJT, Daran JMG, De Ridder D. De novo detection of copy number variation by co-assembly. Bioinformatics. 2012;28(24):3195–202.

53. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

54. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11).

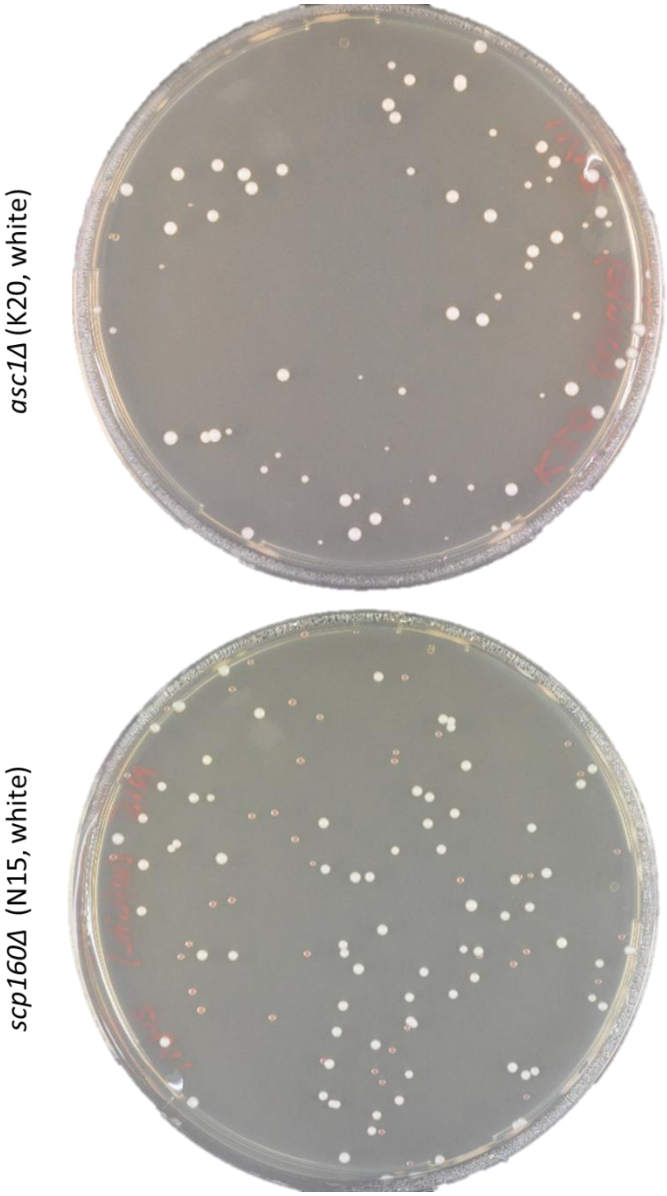
55. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92.

Supplementary information to Chapter 2

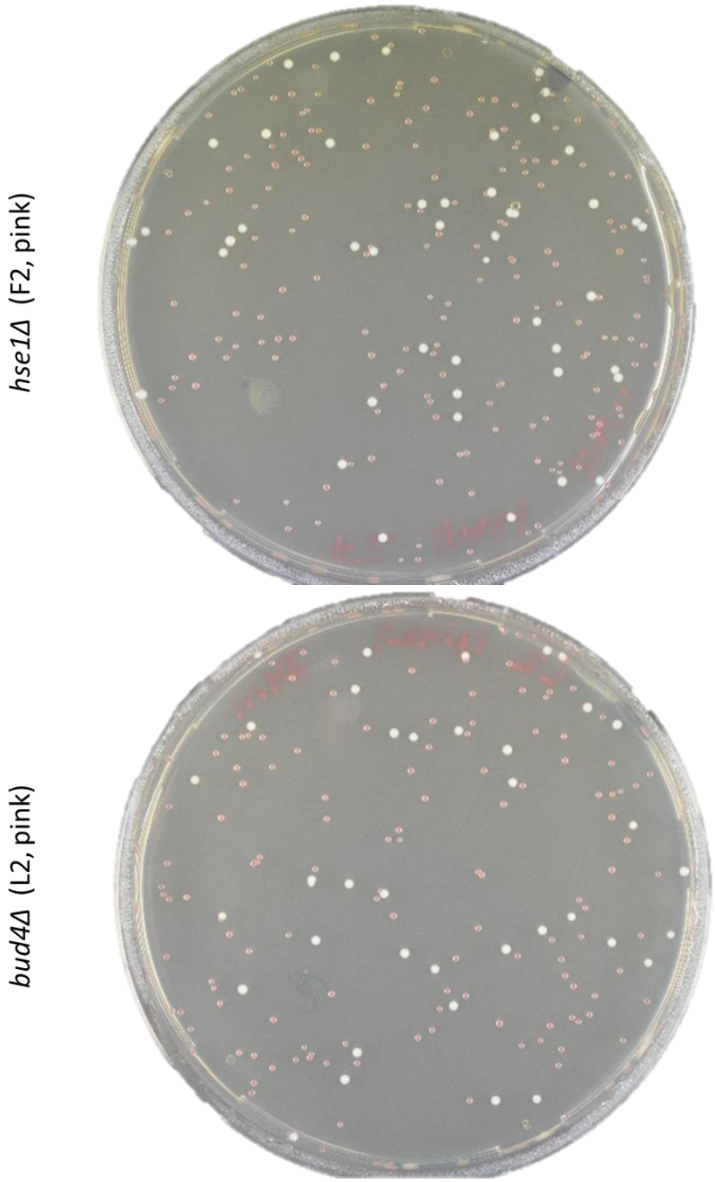
Plate$_{384}$ 9

Plate 33	Plate 34	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	
Plate 35	Plate 36	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	
A	A	HMI1	YMR316C-A	COQ3	YMR316C-B	YOL098C	DIA1	YOL099C	YMR317W	PKH2	ADH6	IZH4	FET4	ITR2	YMR320W	NDJ1	RAD50	WSC3	MRPL17	YOL106W	PRM1	YOL107W	TOS6	INO4	PHA2	A
A	A	CSR2	ZAP1	SRO7	YJL055W	VMA13	PEP8	IRC16	TDH1	YPR039W	IRC8	TIP41	YJL049W	PUF2	UBX6	OPI11	RTT101	YPR045C	AIM22	MCM16	YJL045W	MSF1	GYP6	ATG11	YJL043W	B
B	B	ZEO1	CMK2	SHR5	RPS7A	MDY2	MCH5	MSB4	FAA1	SKM1	COX11	YOL114C	ODC1	PAP2	DAK1	MSN1	COG8	RRI2	FPR3	YOL118C	YML090W	MCH4	GIM5	RPS19A	RAD10	C
B	B	MAK3	LOH1	NHP6A	IRC18	YPR053C	SNX4	SMK1	MAD2	BRR1	UBC8	YMC1	YPT31	YPR059C	SPO73	ARO7	THO1	JID1	RRT13	FCY1	BUB1	YPR063C	YGR201C	ROX1	ADE3	D
C	C	SMF1	YML095C-A	TRM11	YML096W	MDH2	VPS9	YGK3	ARG81	VPS68	TSL1	YOL131W	YML100W-A	GAS4	CUE4	PFK27	YML102C-A	0	CAC2	BSC6	0	RTC1	NUP188	YBR232C	MDM1	E
C	C	UBA3	HSE1	HOS1	PRS3	SPE3	YHL039W	MED1	YHR003C	YPR071W	NEM1	LTP1	STP2	TKL1	SOD2	OPY2	YHR009C	YPR076W	THR1	YPR077C	PPA1	0	SRB2	MRL1	0	F
D	D	DYN2	URA5	GLC3	PML39	YER064C	YML108W	YER077C	ZDS2	ICP55	DAT1	DOT6	ATR1	TRP2	NAB6	MET6	YML117W-A	IES5	NGL3	AIM11	YML119W	RAD51	NDI1	SHC1	GTR1	G
D	D	YPR084W	FYV4	VPS69	HTD2	YPR089W	YHR127W	YPR090W	YHR131C	YPR092W	YHR180W	ASR1	PFS1	SYT1	MDM31	YPR096C	YLL007C	YPR097W	RIM13	YPR098C	GOR1	YPR099C	PPM2	MRPL51	YOL150C	H
E	E	YER097W	YML122C	UBP9	PHO84	CAF130	TUB3	YGR210C	MSC1	RIM4	YML131W	SNF6	MVP1	YSC84	SNO1	YSC83	SNZ1	YHR032W	YMR099C	YHR045W	MUB1	SSZ1	SRT1	YHR140W	YMR102C	I
E	E	SNT309	ENB1	SPT10	YOL159C	NIT2	YOL160W	LSM1	YOL162W	MTC1	YOL163W	ALB1	OPI3	YJL120W	HOC1	RPE1	BNA2	YJL118W	YJR079W	YJL119C	EAF6	PHO86	ACF4	NCA3	EMC2	J
F	F	YHR162W	YMR103C	MTG2	PGM2	SVP26	YKU80	CTF8	SPG4	EGD2	MYO5	RRT7	HFD1	YLL044W	YMR111C	YBT1	YMR114C	LDB18	MGR3	YLL059C	ASC1	YLR030W	YMR119W-A	YLR031W	ADE17	K
F	F	ASF1	BUD4	MDV1	VPS25	GZF3	URA8	PRM10	ADO1	YJL107C	ABM1	IME2	YMR1	MEF2	YJR111C	LSB6	YJR115W	CHS6	RSF2	SAP185	YJR128W	MRPL49	YJR129C	BCK1	STR2	L
G	G	RAD5	RPL15B	SMF3	YMR122C	MLH2	PKR1	YLR036C	YMR124W	PAU23	STO1	COX12	DLT1	RIC1	SAS2	YLR040C	POM152	YLR041W	YMR130W	YLR050C	JLP2	IES3	REC114	RPL13B	GID8	M
G	G	TOK1	MCM22	SRS2	MET5	SIP4	YJR146W	ARG3	HMS2	ALY2	YJR149W	TAX4	DAL5	IML2	YJR154W	SCP160	OMA1	PRY1	TVP38	ICS3	TGL4	JEM1	PXL1	ARG2	SRL3	N
H	H	MRPS8	ICL2	0	0	EAR1	SUT2	YMR181C	YPR012W	YMR209C	YPR014C	URA10	YPR015C	YMR279C	DSS4	FKS3	RLF2	GLC8	ATP20	ELP6	YPR027C	TGL3	YOP1	YMR315W	APL4	O
H	H	YJL068C	SRP40	0	0	YJL067W	PTR2	MPM1	PCK1	YJL064W	UBP11	DLS1	BAS1	MRPL8	SKG1	LAS21	SIR1	BNA3	NFT1	YHC3	YKR104W	BIT61	VBA5	IKS1	COX19	P
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	

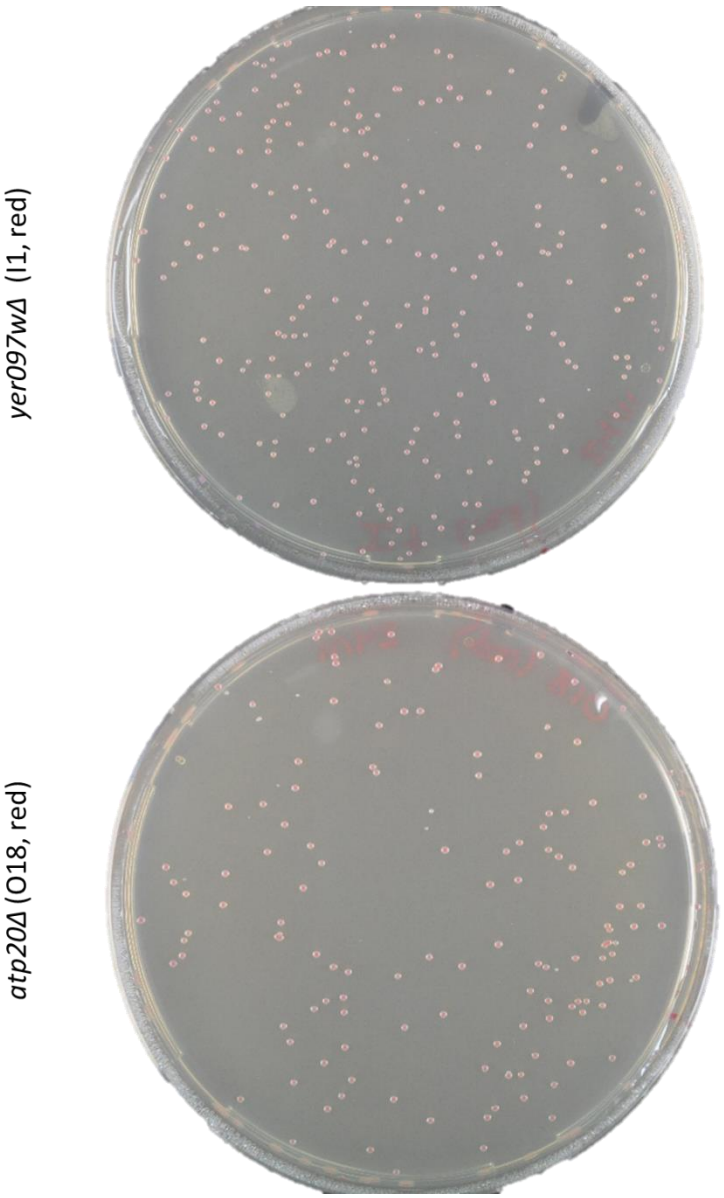
Supplementary Figure 1. Positions of mutant strains on Plate 9 of the yeast deletion library when in a 384 density format.



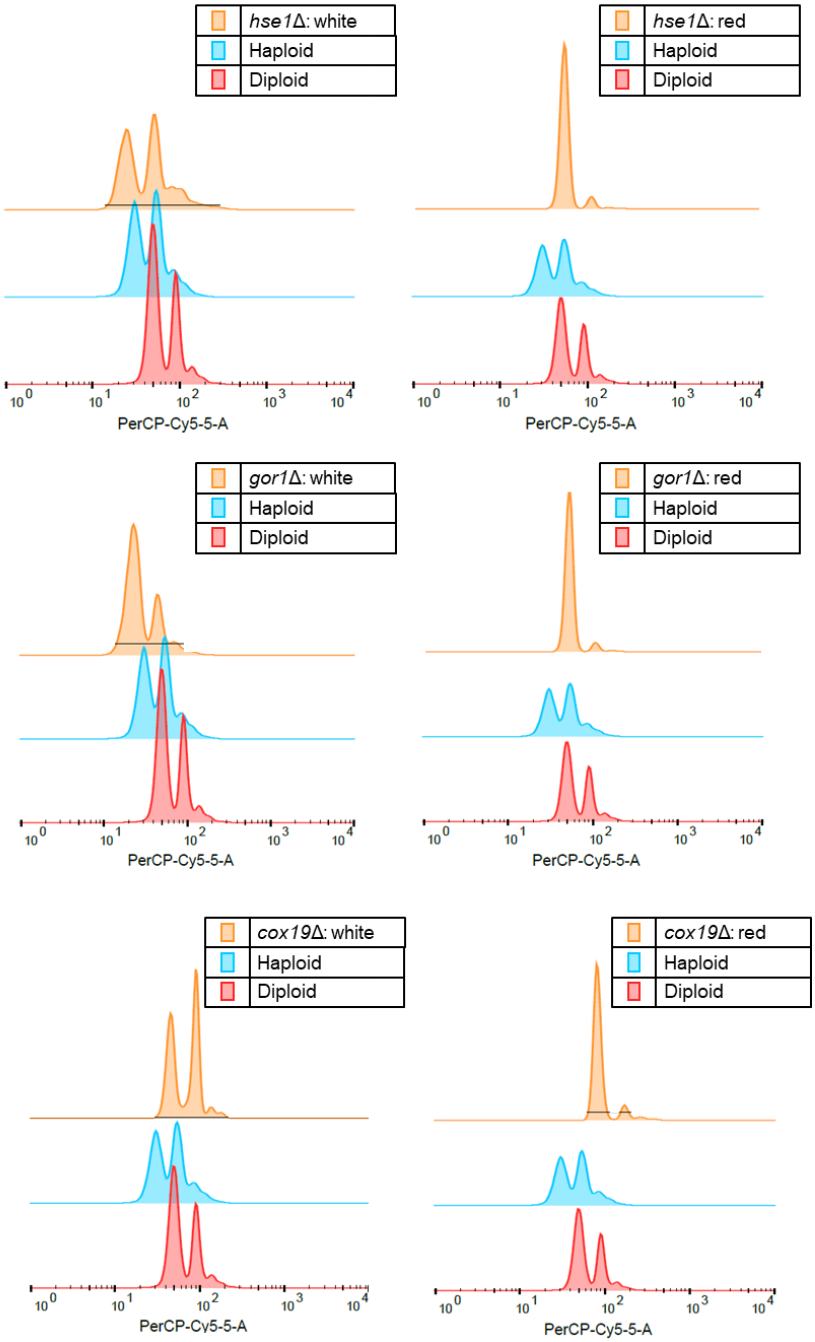
Supplementary Figure 2. Single colonies of the two strains *asc1Δ* and *scp160Δ* mated to donor strain *Sc*_HO19 and put through the replica plating procedure outlined in Figure 5C.



Supplementary Figure 3. Single colonies of the two strains *hse1*Δ and *bud4*Δ mated to donor strain *Sc*_HO19 and put through the replica plating procedure outlined in Figure 5C.



Supplementary Figure 4. Single colonies of the two strains *yer097w*Δ and *atp20*Δ mated to donor strain *Sc*_HO19 and put through the replica plating procedure outlined in Figure 5C.



Supplementary Figure 5. Assessment of ploidy level of CRI-SPA generated strains. Strains originate from the final plate of an alternative plating procedure in which diploids were grown on raffinose before transfer onto galactose. The strains were streaked out into single colonies and in each case, a white and a red single colony of the respective strain was analysed. Cells were stained with propidium iodide and the histograms show the distribution of cells in accordance to their relative DNA content when analysed by flow cytometry. Ploidy was assessed by comparing histrograms of each clone to that of a known haploid and diploid reference strain.

Supplementary Table 1. Primers used in this study. USER cloning tails are indicated in bold; gRNA protospacer sequences are underlined; long non-binding tails are written in lowercase letters.

Primer name	Sequence (5' to 3')	Purpose
Primers for cloning		
HOP89	ACACGCGAUGCTGGAGCTCATAGCTTCA	Construction of Cas9 plasmid pHO8
HOP90	ACGTGCGAUCTATAGGGCGAATTGGGTAC	Construction of Cas9 plasmid pHO8
HOP91	ATCGCGTGUGAGCTCGCTGAGGACTT	Construction of Cas9 plasmid pHO8
HOP92	ATCGCACGUCAGCTGAAGCTTCGTACG	Construction of Cas9 plasmid pHO8
HOP60	GTTTTAGAGCUAGAAATAGCAAG	Linearizing plasmid pCfB3050 for cloning of new gRNA target sequence
HOP61	GATCATTTAUCTTTCACTGC	Linearizing plasmid pCfB3050 for cloning of new gRNA target sequence
HOP173	ATTCATGTGATAATTTGACGGTTTTAGAGCT	Cloning gRNA target sequence for downstream *ADE2*, (+) strand
HOP174	CGTCAAATTATCACATGAATGATCATTTAT	Cloning gRNA target sequence for downstream *ADE2*, (-) strand
HOP190	AATTGTAGAGACTATCCACAGTTTTAGAGCT	Cloning gRNA target sequence for *ADE2*, (+) strand
HOP191	TGTGGATAGTCTCTACAATTGATCATTTAT	Cloning gRNA target sequence for ADE2, (-) strand
Primers for sequencing		
HOP48	GCGTGTACGCATGTAAC	Sequencing Cas9 plasmid pHO8
HOP84	TCGAGATCGCCGAATGGC	Sequencing Cas9 plasmid pHO8
HOP93	AATTCAAAGTTCTGGGCAATACCG	Sequencing Cas9 plasmid pHO8
HOP94	TGGCCGAAGATGCCAAGCTTCAAC	Sequencing Cas9 plasmid pHO8
HOP95	CAGTCCTTCATCGAAAGGATGAC	Sequencing Cas9 plasmid pHO8
HOP96	TGCAGACCGTTAAGGTCGTGGATG	Sequencing Cas9 plasmid pHO8
HOP97	ATTACTCTGAAGTCTAAGCTGGTC	Sequencing Cas9 plasmid pHO8
HOP98	AAAGACCTCATCATTAAGCTTCCC	Sequencing Cas9 plasmid pHO8
Primers for strain construction and verification		
HOP64	tataacaatcaagaaaaacaagaaaatcggacaaaacaatcaagt**GCATAGGCCACTAGTGG**	Amplification of hphNT1 from pMEL12 for integration at *ADE2* locus
HOP65	attttataattatttgctgtacaagtatatcaataaacttatataCATCAAT**AGGCACCTTCG**	Amplification of hphNT1 from pMEL12 for integration at *ADE2* locus
HOP175	tgatactgttctcataaacatgtgactgcattggtggttgGGGTTCTC**GAGAGCTCG**	Amplification of *Kl URA3* from pCfB390 for integration downstream *ADE2*
HOP176	tccgagggtaactgtgatagcttcaaaagacttttagcatGGTCTAG**AGATCCCAATACA**	Amplification of *Kl URA3* from pCfB390 for integration downstream *ADE2*
HOP87	GATGCAAGTGAAGGTTGTCCGATAAC	Verifying integration of Cas9 at X-3
HOP99	TGCTGTGGCGATCGGTATTG	Verifying integration of Cas9 at X-3
HOP118	CTTGCTCTTGGCGGCTACTG	Verifying integration of Cas9 at X-3
HOP119	GAGGGTAATGTCTCTGAAGCTGC	Verifying integration of Cas9 at X-3
HOP52	GACAATTCAACGCGTCTG	Verifying integration of hphNT1 at the *ADE2* locus
HOP76	GATGGCTGTGTAGAAGTACTCG	Verifying integration of hphNT1 at the *ADE2* locus
HOP81	TGGATAGTCTCTACAATTGG	Verifying integration of hphNT1 at the *ADE2* locus
HOP140	TCGTTGGATCTCTCTTCTAAGTACATCC	Verifying integration of hphNT1 at the *ADE2* locus
HOP141	AGCGTTGATTTCTATGTATGAAGTCCAC	Verifying integration of hphNT1 at the *ADE2* locus
HOP187	GTTTCGATGCAACCGGACTTGC	Verifying integration of *Kl URA3* dw *ADE2*
HOP188	CGAGAGGTGACGTGAAATTCTTCC	Verifying integration of *Kl URA3* dw *ADE2*
HOP189	CTACGGCACTTGAATTAGACTGCC	Verifying integration of *Kl URA3* dw *ADE2*
HOP196	ACGATATGGGAGGAAGAGAAGAAGG	Verifying integration of *Kl URA3* dw *ADE2*

CHAPTER 3: **High-throughput, CRI-SPA mediated transfer and integration of a biosynthetic pathway into multimember strain libraries – Employing the yeast deletion collection for metabolic engineering of the aromatic amino acid pathway in yeast**

Abstract

Obtaining adequate titers, rates and yields is a central goal in the metabolic engineering of microbial cell factories. However, prediction of phenotypes from genotypes remains a challenging task, making the development of cell factories a slow and costly process. Due to complex biology and limited understanding of metabolism and regulatory mechanisms, non-targeted, high-throughput approaches operating at a genome-scale level can be a powerful alternative to rational metabolic engineering strategies. This study demonstrates how the novel CRI-SPA platform can be used together with the genome-wide yeast deletion collection to screen for mutations improving the performance of a yeast cell factory. We applied this method to optimise yeast-based production of betaxanthin pigment, which functioned as a proxy molecule for medicinally important benzylisoquinoline alkaloids derived from the aromatic amino acid pathway. Using CRI-SPA, betaxanthin synthesising genes were transferred to the mutant strains of the yeast deletion collection in less than a week. Mutations with strong effects on pigment production were identified and by reverse engineering, improved intracellular concentrations of betaxanthin was confirmed for a *cse2*Δ and *yor1*Δ mutant, which were recognised as top hits in the initial screening. Furthermore, systematic enrichment analysis was carried on the generated gene sets, identifying overrepresented Gene Ontology (GO) terms amongst top-scoring mutants. This study demonstrates the usefulness of CRI-SPA in cell factory optimisation, showing that the yeast deletion collection can readily serve as a screening platform for metabolic engineering designs, and additionally setting the stage for wider applications of the method.

Introduction

In order to decrease CO_2 emissions and limit global warming, cost-effective and sustainable processes to replace current oil-based production routes are urgently needed. In industrial biotechnology, engineered microbes such as yeast, filamentous fungi and bacteria are employed to convert a renewable substrate into a product of interest. These microbial systems, often referred to as microbial cell factories, may thus provide sustainable, alternative production routes to the traditional petro-chemical ones. *Saccharomyces cerevisiae* is a popular host for the bio-based production of a large number of industrially relevant compounds, including fuels, chemicals and therapeutic proteins (1). Yeast cell factories are, however, frequently hampered by yields that are significantly lower than the theoretical maximum. This is a problem both in terms of sustainability and profitability, especially when the production of high-volume, low-value bulk chemicals are considered, since competing petro-based production typically offers significantly cheaper production. To optimise the performance of the cell factory, a metabolic engineering strategy is typically applied in which the native metabolism of the producing organism is re-routed in order to divert metabolic fluxes towards production of the desired compound. Nevertheless, owing to the inherent complexity of biological systems, accurate prediction of phenotypes from genotypes remains a challenge and the development of a new microbial cell factory is both a costly and time-consuming endeavour.

An estimated 50 million USD is required to generate a production strain for a new bioprocess (2) and, as an illustrative example, it was calculated to have taken 130 R&D person years to develop the yeast-based semi-synthetic production of the antimalarial drug artemisinin (3) (4).

For the reasons outlined above, non-targeted, non-rational screening approaches at the genome-scale can be a viable option for discovering new improved designs for cell factories. The genome-wide deletion collection of yeast (5,6) has been a valuable resource for functional genomics studies and cellular network discovery (7). However, although the collection has been used in many studies screening for tolerance to inhibitors commonly found in industrial fermentations - examples including acetic acid (8), vanillin (9) and ethanol (10) - the collection has only rarely been used to screen for improved production of the intended fermentation product. In 2011, a study by Suzuki *et al.*, exploited the deletion collection to screen for genes involved in gluthathione production (11) and in 2012, the same group also screened the collection to find mutations leading to increased production of organic acids (12). Additionally, a few studies have used the deletion collection to screen for improved production of non-native yeast products, including cellulases (13), L-lactate (14) and carotenoids (15). The reason behind the relatively limited number of studies exploiting the collection for metabolic engineering purposes may be simple, as it is an obvious challenge to quantify production of a target molecule in a large library of strains. Moreover, when screening for a heterologous product, each strain in the deletion library needs to be transformed with exogenous DNA, hence requiring methods for high-throughput yeast transformation.

CRI-SPA is a mating-based, CRISPR-Cas9 assisted method for high-throughput, parallel manipulation of the yeast deletion mutants. We have previously shown that CRI-SPA can efficiently introduce a second genetic deletion in the strains of the deletion collection, enabling genetic interaction screens (Chapter 2 of this thesis). Here, we show that CRI-SPA can also be used as a novel method for facilitating easy introduction of a biosynthetic pathway to the yeast deletion library. Aiming to demonstrate the usefulness of CRI-SPA-based genome-wide screens in cell factory development, we applied CRI-SPA to the optimisation of yeast-based production of aromatic compounds by adopting a biosensor approach in which we transfer genes for L-tyrosine-derived synthesis of betaxanthin to the ~4800 strains of the deletion library. By subsequent analysis we identified and verified mutations that caused increased intracellular levels of betaxanthin in yeast and that would be interesting future engineering targets for increased production of L-DOPA - a key precursor in the production of benzylisoquinoline alkaloids such as codeine and morphine. Furthermore, enrichment analyses were applied to the results from the genome-wide screen in order to find Gene Ontology (GO) terms that were over-represented amongst top-scoring mutants. This study establishes the usefulness of CRI-SPA in cell factory optimisation as it enables the deletion collection to readily serve as a screening platform for metabolic engineering designs.

Background

The importance of the aromatic amino acid pathway in biotechnology

A biosynthetic pathway which has been the focus of many metabolic engineering studies in yeast is the aromatic amino acid pathway, as numerous industrially and medicinally important compounds can be derived from its intermediate and final products (16). The pathway, outlined in a simplified manner in Figure 1, starts with the condensation of the two

precursor molecules phosphoenolpyruvate (PEP) (derived from glycolysis) and erythrose-4-phosphate (E4P) (derived from the pentose phosphate pathway) to make the aromatic molecule 3-deoxy-D-arabinoheptulosonate (DAHP). The reaction is catalysed by isoenzymes Aro3 and Aro4, and in a series of downstream reactions, referred to collectively as the Shikimate pathway, penta-functional enzyme Aro1 converts DAHP to shikimic acid and then further to chorismic acid by additional action of Aro2. From this point, the pathway branches out to synthesize the aromatic amino acids L-tryptophan (L-Trp) and L-tyrosine (L-Tyr) on one side, and L-phenylalanine (L-Phe) on the other.

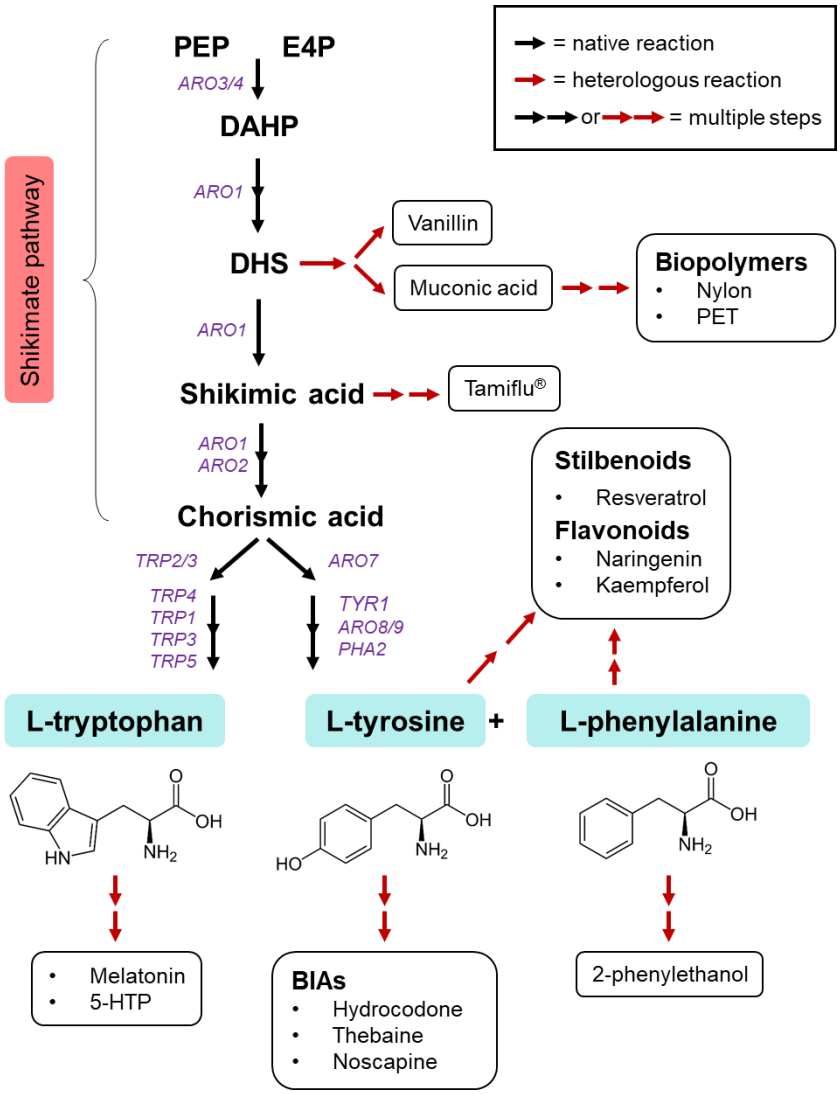


Figure 9. Simplified representation of the aromatic amino acid pathway in *Saccharomyces cerevisiae* and illustrative examples of heterologous compounds produced by yeast cell factories for applications in industry, food, health and medicine. PEP - phosphoenolpyruvate (PEP); E4P - erythrose-4-phosphate; DAHP - 3-deoxy-D-arabinoheptulosonate; DHS – dehydroshikimate; PET – polyethylene terephthalate; 5-HTP - 5-hydroxytryptophan; BIAs – benzylisoquinoline alkaloids.

Aromatic molecules are an important class of compounds for the chemical industry, where they serve as building blocks in the synthesis of many other chemicals. However, their production is typically dependent on extraction and synthesis from raw oil and involves the use of health damaging and environmentally polluting chemicals. Furthermore, aromatic molecules are important to the pharmaceutical and nutraceutical industry as many of these compounds have therapeutic and/or antioxidant properties, and to the flavour and fragrance industry due to their pleasant taste and smell. In these cases, the molecules are often of natural origin, produced as secondary metabolites by plants. Although the production of natural compounds is a billion-dollar industry (17–19), extraction from the native plant hosts can be problematic due to low yields (production and extraction), resulting in high demands for arable land for cultivation with many corresponding environmental impacts. Concomitantly, chemical synthesis of these molecules may be difficult due to their complex chemical structures.

Microbial cell factories represent an attractive alternative to both petro-chemical and plant-based production of aromatic molecules and a great number of industrially and pharmaceutically relevant compounds can be derived from the aromatic amino acid pathway in several microbial host organisms, of which *Escherichia coli* and *S. cerevisiae* are the most frequently used (16). One important advantage of *S. cerevisiae* over *E. coli* is the ability of the organism to express membrane-bound enzymes such as cytochrome P450 oxygenases (20,21), which are key for catalysing the formation of new carbon-oxygen bonds in the synthesis of many plant-derived, aromatic compounds. Production of a number of aromatic molecules of industrial value is now possible in yeast, either by native host metabolism or by expression of heterologous genes from plants and other organisms. Several examples of these compounds are illustrated in Figure 1 and are further discussed below.

Aromatic compounds produced by yeast cell factories can be derived from different branches of the aromatic amino acid pathway. Upstream products derived from the Shikimate pathway include muconic acid (22), which is a precursor used for the production of bio-based nylon (23) and polyethylene terephthalate (PET) (24); vanillin (25), which is the most commonly used flavouring agent in food industry (26); and shikimic acid (27), which is used in the production of anti-influenza drug Tamiflu® (28). Moving downstream of the pathway, the L-Trp branch is used for the production of the important medicinal compounds melatonin (29) and the serotonin precursor 5-hydroxytryptophan (5-HTP) (30); whereas the L-Phe branch has been used for the production of the widely popular scenting agent 2-phenylethanol (2-PE) (31). Moreover, the L-Phe and L-Tyr branch are used in the derivation of several secondary metabolites belonging to the family of flavonoids and stilbenoids, which have multiple applications in human health due to their antioxidant, antimicrobial and anticancer properties. Examples of yeast based production of these molecules include naringenin (32), kaempferol (33) and resveratrol (34). Finally, another important class of compounds derived from the L-Tyr part of the pathway are the benzylisoquinoline alkaloids (BIAs). Many of these molecules have therapeutic effects and include the pain-relieving drugs morphine and codeine, the antimicrobial compounds sanguinarine and berberine (35), and the cough-suppressant, anticancer drug noscapine (36). These molecules have intricate chemical structures and the reported *de novo* synthesis of the BIAs hydrocodone and thebaine - semi-synthetic opium-derivatives – in *S. cerevisiae* in 2015, marked a milestone in yeast synthetic biology, requiring expression of more than 20 heterologous and native yeast genes (37).

As briefly touched on, there is a vast range of important molecules that can be derived from the aromatic amino acid pathway. However, although many of these are specialty chemicals with a high market value, the majority of microbial systems for their production are yet to deliver commercially viable titers, rates or yields (16,38). The aromatic amino acid pathway draws from precursors originating from the primary yeast metabolism and the synthesis of L-Tyr, L-Phe and L-Trp has a heavy energy burden, requiring twice the amount of ATP than the synthesis of most other amino acids (39). The pathway is therefore under strict regulation on several levels in the cell – allosterically, transcriptionally and translationally (40). For these reasons, metabolic engineering of the pathway is complex and rational engineering designs intended to improve production of compounds derived from the pathway may have minor effects. In this scenario, global approaches addressing the problem at a genome-scale level might be necessary to improve the performance of the cell factory and to identify new strategies for informed and targeted metabolic engineering.

A high-throughput screening approach for improved L-DOPA production

Genome-wide studies necessitates the use of high-throughput methods in order to screen a large number of genetic designs for their impact on the investigated phenotype. In a 2015 study by DeLoache *et al.*, an enzyme-based biosensor for L-3,4-dihydroxyphenylalanine (L-DOPA or levodopa) was developed (41), enabling high-throughput screening for its production. L-DOPA can be directly derived from L-Tyr in yeast by action of the CYP76AD1 enzyme, a cytochrome P450 from the sugar beet *Beta vulgaris*. This constitutes the first committed step towards production of BIAs in yeast, as L-DOPA can be further converted to the important branch point intermediate (S)-reticuline, from which sanguinarine, berberine, and opioids such as thebaine are all derived. Moreover, L-DOPA is itself an important molecule as it is used in the clinical treatment of Parkinson's disease (42).

The design of the L-DOPA biosensor developed in the work by DeLoache *et al.* (2015) is illustrated in Figure 2. The biosensor makes use of the enzyme DOPA dioxygenase (DOD) from the flowering plant *Mirabilis jalapa* to convert L-DOPA into betalamic acid, which, by spontaneous reactions with freely available amines, forms a number of molecules collectively referred to as betaxanthins. These are pigments of intense yellow colour and which are highly fluorescent. Betaxanthin may thus serve as a proxy for medicinally important L-DOPA and other aromatic BIAs derived from the aromatic amino acid pathway, while possessing optical properties that allows HT screening using either colorimetric or fluorescent assays. For these reasons, betaxanthin production in yeast offered an ideal test system for the CRI-SPA-based, genome-wide screens conducted in this study as it enabled facile screening and identification of potential engineering targets for improved L-DOPA production amongst the yeast deletion library.

Results

Construction of a Universal Donor Strain for transferring pathways into a well-defined chromosomal integration site

In our previous study using CRI-SPA, a disrupted allele was transferred from its native chromosomal locus in the Universal Donor Strain (UDS) into the same (but non-disrupted) locus of the recipient strain (Chapter 2 of this thesis). In this study, the CRI-SPA system is re-purposed to be used in cell factory optimisation, transferring a biochemical

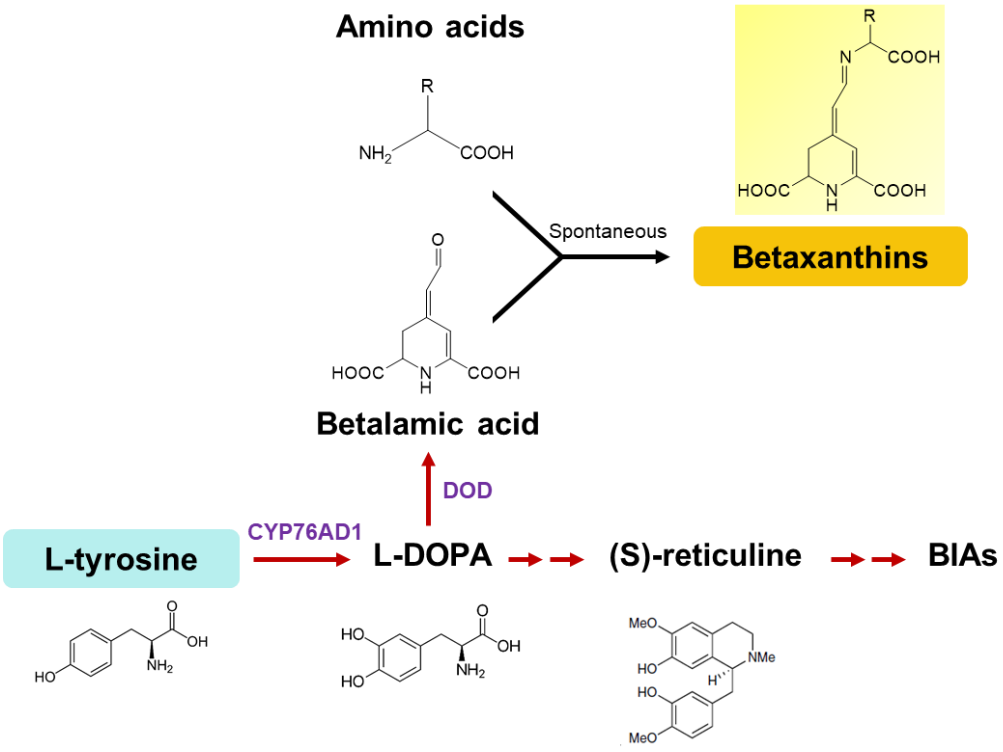


Figure 10. A betaxanthin-based biosensor enabling high-throughput screening for production of L-DOPA in yeast. By expression of the P450-encoding CYP76AD1 gene from *Beta vulgaris*, intracellular L-tyrosine in yeast may be converted to L-DOPA, which is the first comitted intermediate in the synthesis of (S)-reticuline and downsstream benzylisoquinoline alkaloids (BIAs). To allow high-throughput screens for the production of L-DOPA, a second enzyme, DOPA dioxygenase (DOD) from *Mirabilis jalapa*, is employed to convert the L-DOPA into betalamic acid, which spontaneously reacts with freely available amines in the cell to form a group of highly colourful and fluorescent pigments collectively referred to as betaxanthins.

production pathway to a set of recipient strains, which upon completed transfer can be screened for their production capabilities. This set-up offers flexibility in terms of where in the donor genome the pathway genes should be integrated and subsequently transferred from. Cas9 efficiency is highly dependent on the gRNA used for directing the nuclease to its target sequence (43) and due to chromosomal structures, cleavage efficiency may also vary depending on the targeted genomic region. Aided by results from a previous study investigating Cas9-targeting of several well-characterised chromosomal integration sites in yeast, it was decided that the XII-5 integration site (44) would be a suitable locus for harbouring the biosynthetic pathway to be transferred by the CRI-SPA system, as the site proved efficient both in terms of Cas9-gRNA targeting and capacity of incorporating DNA by homologous recombination (HR) (45).

In a first step of repurposing the CRI-SPA system, a second *URA3* marker was integrated in chromosome XII of the CRI-SPA donor strain *Sc*_HO10, yielding the new strain *Sc*_HO20. In this strain, the XII-5 site was hence enclosed between two *URA3* markers – one adjacent to the centromere and an additional one downstream of the XII-5 site. When mated to a recipient strain, the Cas9-induced double-strand break (DSB) of the recipient chromosome is repaired by HR using the UDS chromosome as repair template. As explained in Chapter 2 and recapitulated here, this HR mediated repair most often occurs by gene conversion only, but occasionally gene conversion is accompanied by a crossover event. In

the latter case, the cut recipient chromosome and the donor chromosome exchange chromosomal DNA ranging from the break site to the telomere. In experiments where the donor and the recipient strain are non-isogenic, this event is undesirable as the selected strain will contain a chimeric chromosome of genetic material from both the donor and the recipient chromosome. However, by inserting a second *URA3* marker distal to the region involved as template for HR-mediated DSB repair, recombinants formed as the product of gene conversion accompanied by crossing-over can be counter-selected on medium containing 5-FOA as the *URA3* marker would be transferred along with the intended DNA into the recipient chromosome. The chromosomal architecture of the *Sc*_HO20 UDS is schematically illustrated in Figure 3 and was used in subsequent CRI-SPA transfers of genetic material harboured within its XII-5 site.

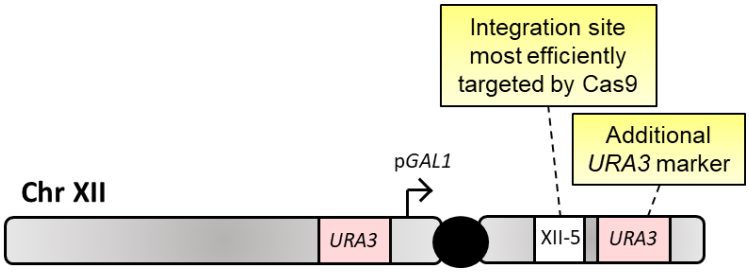


Figure 11. Chromosomal architecture of the CRI-SPA Universal Donor Strain Sc_HO20. The *pGAL1-URA3* cassette inserted at the centromere renders UDS chromosomes conditionally stable and counter-selectable. The XII-5 integration site is efficiently targeted by Cas9 and is used for harbouring DNA to be transferred. Insertion of an additional *URA3* marker downstream of the XII-5 site enables counter-selection of CRI-SPA generated clones in which gene conversion-associated crossing over causes genomic scrambling.

Construction of a Universal Donor Strain for high-throughput transfer of betaxanthin-synthesising genes

BIAs, derived from the L-Tyr branch of the aromatic amino acid pathway, include important medicinal molecules such as codeine and morphine. However, their production is complex and yeast-based cell factories for these and other aromatic molecules typically offer modest yields. Rational metabolic engineering may be challenging as the pathway is under strict, multilevel regulation in the cell and we therefore believed that using CRI-SPA together with the yeast deletion collection would be an attractive alternative for identifying mutations influencing production of BIAs and other aromatic compounds. To investigate this possibility, the betaxanthin biosensor strategy previously outlined was employed, allowing HT screening for production of L-DOPA from L-Tyr – the first committed step towards BIAs. To this end, a Universal Donor Strain was constructed capable of producing betaxanthin by integrating the CYP76AD1 gene together with the DOD gene, and the antibiotic resistance marker *NatMX* at the XII-5 integration site. This yielded strain *Sc*_HO25. Upon integration of the betaxanthin-synthesising genes, Cas9 cleavage of the XII-5 site was eliminated and the strain displayed a clear, yellow phenotype (Figure 4). This donor strain was then used to screen the yeast deletion library for improved production of betaxanthin in order to identify mutations that may ultimately improve production of L-DOPA, while simultaneously using this as test case for demonstrating the usefulness of CRI-SPA in a cell factory optimisation process.

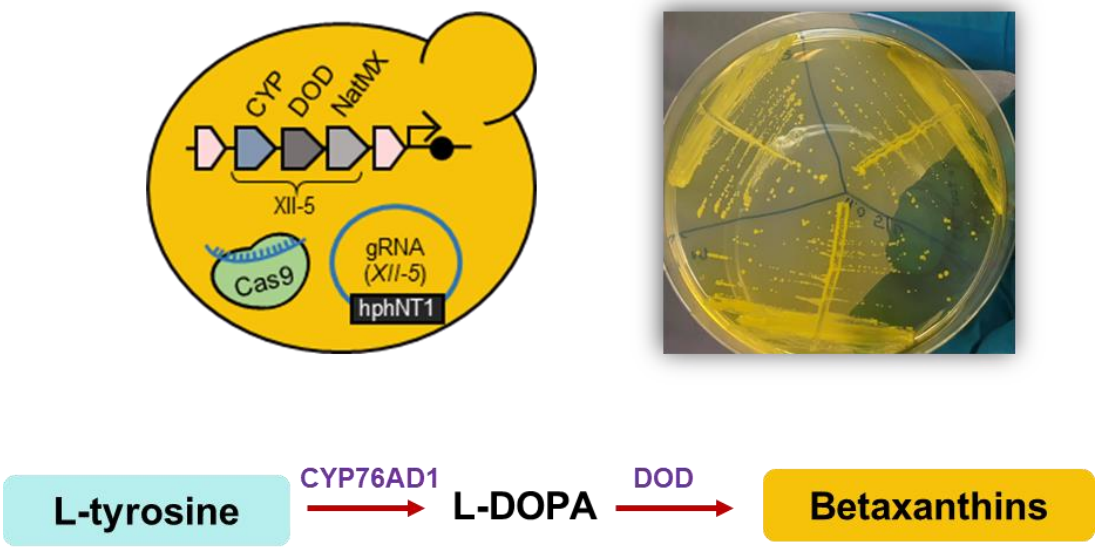


Figure 4. A CRI-SPA Universal Donor Strain for transferring genes for the synthesis of betaxanthin. The CYP76AD1, DOD and *NatMX* selection marker is integrated in the XII-5 site of Sc_HO25 enabling conversion of L-tyrosine to betaxanthin and giving the UDS a visibly yellow phenotype.

Transfer of the betaxanthin pathway to the yeast deletion library and identification of metabolic engineering targets for improved production

The full yeast deletion collection was distributed between 13 rectangular agar plates on which strains were arrayed in a 24 by 16 grid in a 384 density format. Before mating to the UDS, the collection was re-arrayed into a denser, 1536 format (48 by 32 grid), where each mutant strain was present in a 2 by 2 grid of quadruplicates, and where each set of quadruplicates maintained their original position of the 384 density array. The UDS carrying the betaxanthin-synthesising genes and a gRNA plasmid for targeting the XII-5 integration site was then mated to the strains of the deletion collection (Figure 5A) and to transfer the betaxanthin pathway an automated pin replication procedure was used (Figure 5B). Previous experiences from transferring the *ade2* deletion in high throughput to the deletion collection clearly showed the importance of eliminating the growth advantage that a non-targeted, wild-type cell may have over cells that have correctly incorporated the transferred genetic material during the pin replication process. Therefore, extra tyrosine was added to the growth medium of all the plates throughout the procedure in an attempt to lessen potential negative effects of a drained tyrosine pool in high-producing strains, as this amino acid is the main precursor for the synthesis of betaxanthin.

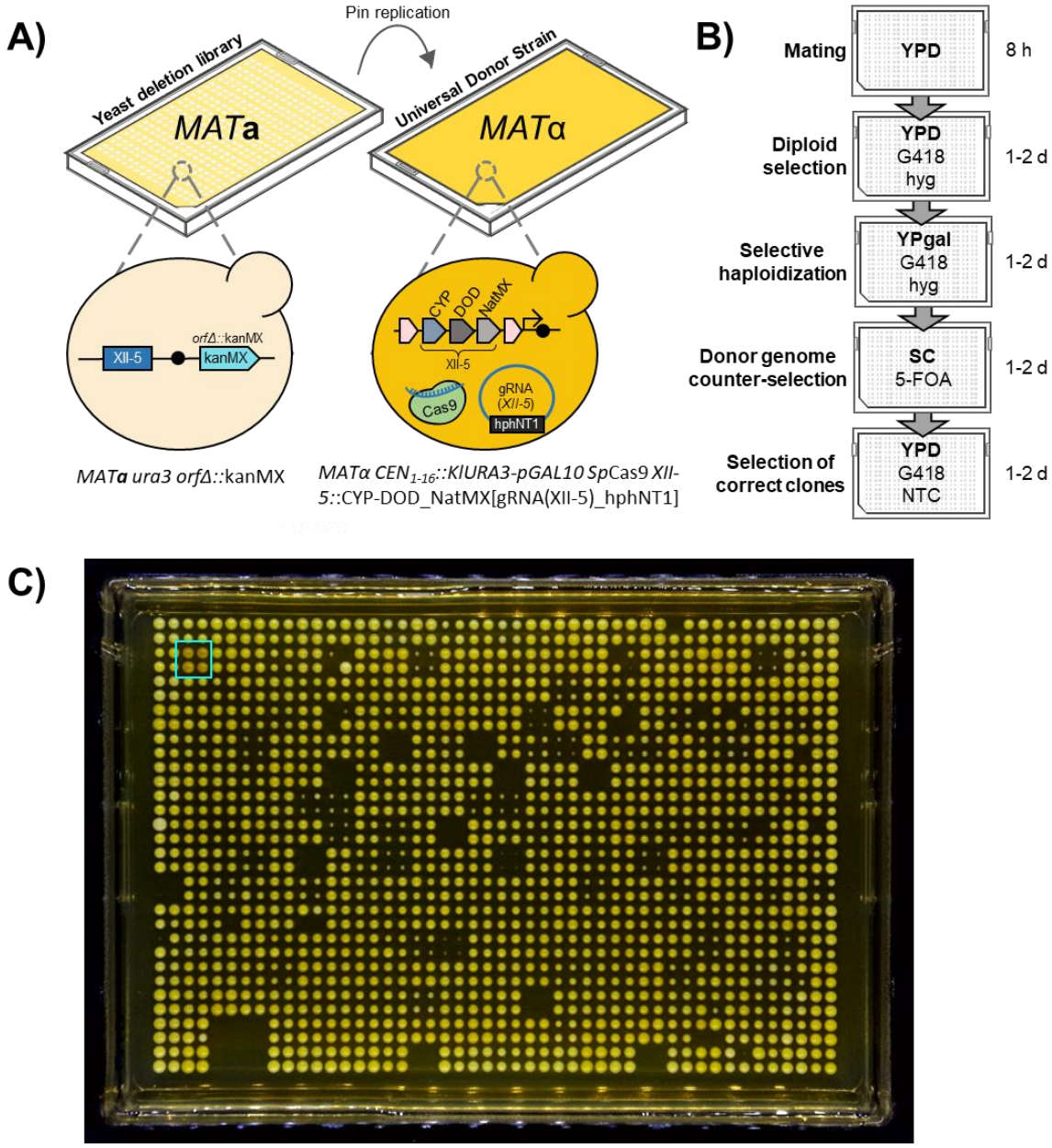


Figure 5. CRI-SPA-mediated transfer of betaxanthin synthesising genes to the yeast deletion library. (A) Experimental set-up. The yeast deletion library was arrayed on rectangular agar plates and were then transferred by replica pinning onto a lawn of the Universal Donor Strain (UDS) Sc_HO25 in order to facilitate mating. The Sc_HO25 UDS harboured the CYP76AD1 and DOD gene needed for L-tyrosine-derived production of betaxanthin and a *NatMX* marker in its XII-5 integration site. The UDS further carried a gRNA plasmid for targeting the native XII-5 site in the recipient genome. (B) Employed pin replication protocol. (C) Final CRI-SPA generated strains on plate 8_{384} (strains are grown in quadruplicates in a 2 by 2 grid). The blue square indicates the position of the *yor1Δ* strain.

Following the transfer and incubation for several days, strains growing on the final selective plates displayed visible differences in the production of the yellow pigment and in some cases, mutants with significantly higher production could easily be identified by eye. This is illustrated in Figure 5C, which shows plate 8_{384} of the experiment (see Supplementary Figure 1 for positions of mutants on this plate) and on which the *yor1*Δ mutant displayed a deep, orange colour. The *YOR1* gene encodes an ATP-binding cassette (ABC) membrane transporter which mediates export of organic anions and several drug molecules (46). In a wild-type, betaxanthin-producing cell, betaxanthin is both retained intracellularly and excreted to the growth medium (41) and it was therefore assumed that the identified *yor1*Δ mutant was accumulating intracellular betaxanthin due to decreased export of the compound. During the work of this study, the same group which developed the betaxanthin biosensor (41), also employed their own system to screen large strain libraries for increased betaxanthin production (47). Their screen likewise identified a *yor1Δ* mutant as a potentially high producing strain and the study confirmed that *yor1Δ* mutants indeed retain a larger fraction of betaxanthin intracellularly than the wild-type. Hence, the observation of the *yor1*Δ mutant in the screen conducted here had a solid biological explanation and was an encouraging quality measure for the method.

After this initial observation, quantification of betaxanthin production was attempted in a more systematic way. Strains were transferred to fresh YPD plates and allowed to grow for 24 hrs before being photographed and analysed using a PhenoBooth Colony Counter. For each strain growing on the plates, a colony size was calculated in terms of area, and a value of colour intensity within that area was obtained. As such we used colour intensity per area as the semi-quantitative metric for betaxanthin production. Strains with significant growth defects and strains with large variation between biological replicates were excluded from the analysis.

To illustrate the range of production levels amongst the generated mutants, the 50 highest and the 50 lowest producers from the screen are presented in Figure 6. The top 150 and bottom 150 mutations are also listed in Supplementary Tables 1 and 2. The highest producer in the screen harboured a deletion of the *CSE2* gene, encoding a subunit of the RNA polymerase II mediator complex that is a key regulator of transcription in eukaryotes (48,49). The *cse2*Δ mutant showed a 6-fold increase in colour intensity compared to the lowest producing strain (*iba57*Δ), and a 2-fold increase over the median colour intensity of the population. The second highest colour intensity (1.9-fold compared to median) was recorded for a strain in which the *MLC2* gene was deleted – encoding the regulatory light chain of the type II myosin Myo1 protein (50) - whereas third highest intensity (1.8-fold compared to median) was observed for the *yor1*Δ mutant previously identified for its strong pigmentation.

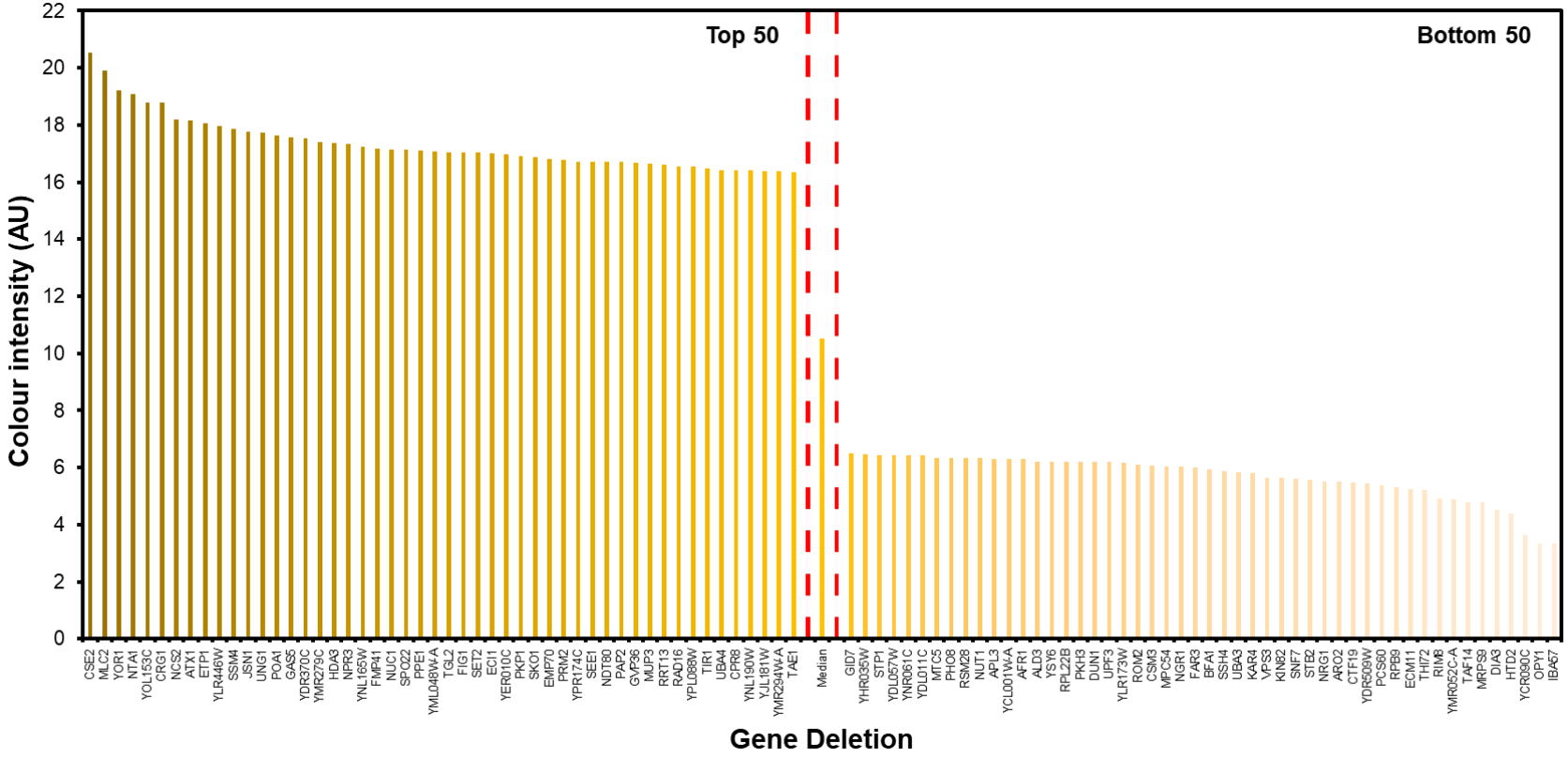


Figure 6. The top 50 highest and bottom 50 lowest colour intensities recorded for mutant strain from the CRI-SPA based transfer of betaxanthin synthesising genes to the yeast deletion library. The median value amongst the mutants is included as a reference.

Analysis of betaxanthin production in reverse engineered strains

In order to assess the validity of the high-throughput screening results, the three mutations yielding the highest production of pigment were reverse engineered into a background strain using conventional genetic engineering methods. First, betaxanthin synthesising genes were integrated in the XII-5 site of BY4741 (the genetic background of the deletion collection) yielding control strain *Sc*_HO26. *CSE2*, *MLC2* and *YOR1* were then individually deleted in this strain by insertion of the *kanMX* cassette at each respective locus, yielding strains *Sc*_HO29 (*cse2*Δ), *Sc*_HO30 (*mlc2*Δ) and *Sc*_HO31 (*yor1*Δ).

Strains, including the non-producing, wild-type BY4741 and the control strain *Sc*_HO26, were grown in liquid and were then analysed by flow cytometry measuring the fluorescent signal (525 – 550 nm) from a population of each strain. In this way, the intracellular levels of betaxanthins in the cells of the different strains could be compared. The average of three biological replicates for each strain is presented in Figure 7. Statistically significant higher fluorescence intensities were observed for the *cse2*Δ and the *yor1*Δ mutants compared to the control strain ($p < 0.05$), with them showing a 1.6 and 1.4 fold increase, respectively. The *mlc2*Δ mutant yielded an average fluorescence that was higher than the control strain (1.1 fold change), but the effect was not statistically significant ($p > 0.05$).

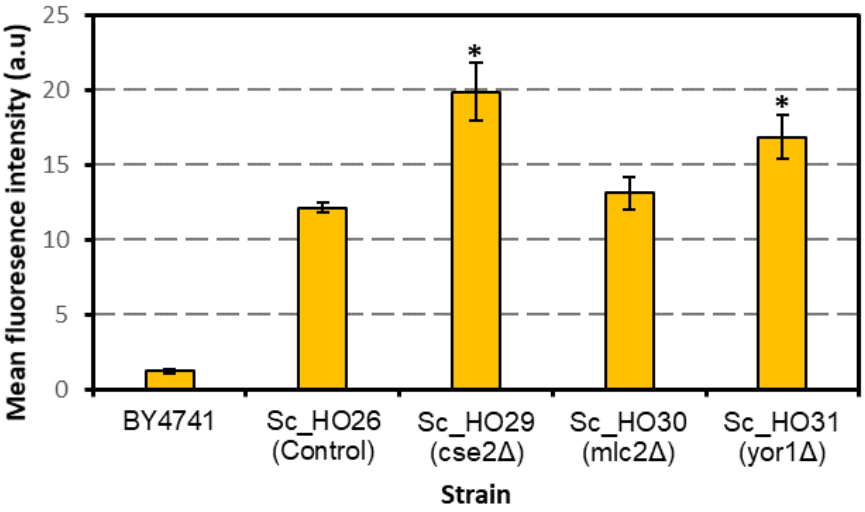


Figure 7. Intracellular mean fluorescent intensities measured by flow cytometry for reverse engineered strains, a betaxanthin-producing control strain and non-producing wild-type BY4741. Asterisks indicate reverse engineered strains with a statistically significant different mean fluorescence intensity versus the control strain ($p < 0.01$).

These flow cytometry results confirmed the positive effect that deletion of *YOR1* has on intracellular retention of betaxanthin, and which had previously also been shown by Savitskaya *et al.* (2019). The fact that this method effectively measures intracellular betaxanthin could explain why the *mlc2*Δ strain did not surpass the control strain, as it is possible that the mutant mainly exports the pigment during cell cultivation. As stated, the *cse2*Δ mutant did, however, demonstrate

significantly higher intracellular fluorescence and may represent a new potential engineering target for increasing betaxanthin production. Although the underlying biological effects of this mutation on the production of betaxanthin and other aromatic compounds were not investigated here, its potential involvement in transcriptional regulation processes is intriguing and is further addressed in the Discussion of this chapter.

Analysis of Gene Ontology (GO) terms amongst top and bottom-scoring mutants

To further extract information from the genome-wide screen, the associated Gene Ontology (GO) terms of the top 150 and bottom 150 mutations were analysed (Supplementary Table 1 and 2). In a first observation, it was noted that only five out of 300 genes were associated with the GO term "metabolic process" (GO:0008152) (Supplementary Table 3) and the great majority of genes were hence connected to other biological functions. We then carried out a GO term enrichment analysis for our two gene lists (top and bottom). After subjecting candidates to the Benjamini-Hochberg test to decrease the rate of false discovery, no significant results were obtained for genes of bottom producers, whereas two GO terms were significantly enriched amongst the top producers ($p < 0.05$). The two terms were both connected to post-translational modifications (PTMs) by methylation and had the following classifications: peptidyl-lysine demethylation (GO:0018027 (biological process) and protein-lysine N-methyltransferase activity (GO:0016279 (molecular function). These terms share the following four genes: *NNT1*, *SEE1 (EFM4)*, *RKM1* and *YBR271W* (*EFM2*) and the latter also includes *YNL042C* (*EFM6*). Information about these genes along with their rankings in the screen is summarised in Table 1.

Table 4. Summary of enriched Gene Ontology (GO) and associated genes amongst the 150 highest-scoring mutants in the betaxanthin screen. Rank indicates the position of each gene in this top list.

Gene name	GO term(s)	Description	Rank
NNT1 (*EFM7*)	GO:0016279; GO:0018027	SAM dependent glycine and lysine methyltransferase; methylates elongation factor eEF1A	113
SEE1 (*EFM4*)	GO:0016279; GO:0018027	Lysine methyltransferase (possibly SAM dependent); methylates elongation factor eEF1A	36
RKM1	GO:0016279; GO:0018027	Lysine methyltransferase; methylates ribosomal subunits proteins L23 and L18	150
YBR271W (*EFM2*)	GO:0016279; GO:0018027	SAM dependent lysine methyltransferase; methylates elongation factor EF2 and EF3	118
YNL024C (*EFM6*)	GO:0016279	Putative SAM dependent lysine methyltransferase, methylates elongation factor eEF1A	78

The GO term-associated genes were all either confirmed or putative methyl transferases that methylate lysine residues in proteins part of the translational machinery. One of them, *RKM1*, targets ribosomal subunits, whereas the other four all target the eukaryotic elongation factor 1A (eEF1A). This highly conserved protein has a main role in delivering

aminoacyl tRNAs to the ribosome for incorporation in extending polypeptide chains, but other functions of the protein are also being discovered (51), where methylation has a key impact. eEF1a is the most extensively methylated protein in yeast and out of only 30 known methyltransferases in the organism, five (Efm1, Efm4, Efm5, Efm6, Efm7) target eEF1A. Interestingly, the screen employed here has thus identified three out of these five methyl methyltransferase along with an additional putative methyltransferase for the same target. Possible implications of this protein on the production of betaxanthin are addressed in the following discussion.

Discussion

CRI-SPA mediated screening of the yeast deletion collection for improved production of aromatic compounds

Microbial cell factories for the production of commodity and speciality compounds offers exciting alternatives to environmentally polluting and climate damaging petro-chemical processes. However, the majority of microbial production systems are not performing at commercially viable levels and in order for these cell factories to be competitive with current, oil-based production, metabolic engineering of their microbial hosts is needed to optimise conversion of substrate into the intended product. Due to complex biology and limited understanding of metabolic and regulatory processes, less targeted screening approaches on the genome scale may offer viable alternatives to rational strategies. In this regard, we recognise that the yeast deletion library is a powerful yet largely unexploited resource and we therefore introduced CRI-SPA as a high-throughput strain construction method to facilitate the use of defined mutant collections in metabolic engineering.

To test the usability of CRI-SPA in cell factory optimisation, we chose to focus on the aromatic amino acid pathway in yeast. This a relevant test case as many important and valuable compounds can be derived from this pathway. Furthermore, the upper part of the pathway is linked to both glycolysis and the pentose phosphate pathway, which means that metabolic engineering strategies may be efficiently counteracted by strict regulatory mechanisms in the cell, preventing redistribution of fluxes. We therefore hypothesised that our genome-wide, non-targeted screening approach using CRI-SPA and the yeast deletion collection could offer advantages as it enables identification of beneficial mutations that are nonobvious from a rational metabolic engineering perspective. In our study we further chose to focus on the production of L-DOPA, which is a key precursor in the synthesis of BIAs. These medicinally important molecules have complex chemical structures and requires expression of numerous heterologous genes often encoding P450s. In this regard, the functionality of the heterologous enzymes outside their native hosts is typically the main bottleneck in the production of BIAs and much attention has been devoted to improving their performance by protein engineering approaches. However, several parameters other than amino acid sequence influence the function of P450s as they require correct expression, folding and co-factor usage in their non-native host (21). Furthermore, they require partnering reductases and to be loaded with an iron-heme prosthetic group. Hence, the cellular environment of the heterologous host is an important factor and in this regard, genome-wide screening of different genetic landscapes may also help the optimisation of P450 dependent production processes.

By transferring a set of betaxanthin synthesising genes via CRI-SPA, we scored the level of production of this compound in the mutant strains of the yeast deletion collection and used the pigment as a proxy read-out for production of L-DOPA. In this study, we chose to focus on deletions which caused increased betaxanthin production in our screen. However, it would have been highly interesting to also investigate genes which upon deletion severely reduced pigment production, as these are logical targets for overexpression. Out of the top three highest scoring mutations, we could verify the phenotype of two by reverse engineering, as strains carrying these mutations produced significantly higher levels of intracellular fluorescence than the control strain. One of these mutations, *yor1*Δ, was also identified in a study by Savitskaya *et al.* (2019) (further discussed below) and leads to increased cellular retention of betaxanthin, which explains its high-scoring position in the screen. It would, however, be interesting to also investigate if this mutation has a similar retention effect on L-DOPA as this could have subsequent positive effects in the synthesis of downstream metabolites in the BIA production pathway.

The second verified mutation in our screen, *cse2Δ*, has a less obvious connection to the phenotype. *CSE2* (or *MED9*) encodes a component of the RNA polymerase II mediator – a multi-subunit complex, which in yeast is composed of 25 different proteins and that has an essential role in transcriptional regulation in eukaryotes (49). Early studies showed that deletion of *CSE2* in yeast affects the expression of a relatively small number of genes (384), with the majority being upregulated (208) (52). The study could not confirm any direct binding of transcriptional repressors or activators to the Cse2 subunit, but it is known that the mediator complex interacts with Gcn4 (53), which is the main transcriptional activator for genes of the aromatic amino acid pathway (40). However, the mediator complex also interacts with a large number of other regulatory proteins in the cell (54). At this point it is therefore difficult to set up a strong hypothesis regarding underlying biological modes through which the *cse2*Δ mutation increases intracellular concentrations of betaxanthin, and it is not yet investigated if this in turn correlates to increased production of the compound. To complicate the matter further, mutation of the *CSE2* gene in two other studies showed elevated frequencies of chromosome mis-segregation (55,56). It could not be discounted that this may influence betaxanthin synthesis simply as a result of variations in chromosome copy number (and subsequently contigs for the inserted pathway), although we would have expected to observe larger deviations between biological replicates during our screening if this indeed was the case. The underlying biological mechanisms for the positive effect on betaxanthin production in the *cse2*Δ mutant should thus be examined further and it would be interesting to investigate the transcriptional profile of the strain.

The CRI-SPA method allowed scoring of phenotypes of each individual strain in the screen, which enabled us to carry out systematic analyses of a set of genes in top and bottom scoring mutants. Looking into the GO terms of the top 150 high-producing, and the bottom 150 low-producing mutations, only five genes were connected to an actual metabolic process. Furthermore, the only significantly enriched GO terms in the two gene sets were both linked to posttranslational modifications of translational machinery proteins. The nature of genes in our top and bottom scoring sets may reflect the strict metabolic regulation of the aromatic amino acid pathway and our non-targeted approach has indeed identified genes that would seem non-intuitive from a metabolic engineering perspective, and which effects on phenotype could not be predicted by genome-scale metabolic modelling. Interestingly, four out of five genes associated with the enriched GO terms are methyltransferase that methylate lysine residues of the eukaryotic elongation factor 1A (eEF1A). This is

the most heavily methylated protein in yeast and it is interesting that our screen has identified three out of in total five known methyltransferases targeting eEF1A in yeast, plus an additional putative methyltransferase of the same target. At this stage, this an intriguing observation, although it is difficult to draw conclusions towards its biological role in the observed phenotype. The function of non-histone protein lysine methylations has only recently begun to be elucidated and the role of methylation in the functionality of eEF1A and other proteins of the transcriptional apparatus is still unclear (57). *In vitro* studies of a yeast eEF1A protein in which methylation sites had been disrupted showed no obvious impairment in translational activities and thus hinted at functions other than translational ones of the methylated protein (58). Later studies investigated the effect of deleting individual genes encoding methyltransferases and observed diverse phenotypes for example linking eEF1A methylation to cytoskeletal interactions and vesicle transport (59,60), to the ability of different viruses to replicate inside the yeast cell (61,62), and to the replicative life span of the cell (63). A recent study also thoroughly characterised a mutant yeast strain lacking all eEF1A methyltransferases and observed increased sensitivity of the mutant to rapamycin and caffeine, possibly indicating a disrupted TORC1 pathway (64). TORC1 is a major signalling pathway in eukaryotes, involved in several cellular activities and responses to stress, including amino acid starvation. In yeast, crosstalk between the TORC1 and the general amino acid control (GAAC) signalling pathway has been shown to regulate translation of *GCN4* mRNA (65). Gcn4 is the major transcriptional activator of genes in the aromatic amino acid pathway and, as addressed above, it also interacts with the RNA polymerase II mediator complex of which *CSE2* encodes a subunit of. It is possible that mutations to the methyl transferase genes identified in the GO term analysis have broad effects on the metabolic landscape of the cell, and that pleiotropic phenotypes resulting from their deletion could benefit betaxanthin production in a way that is not straight forward to elucidate but which would be interesting to further investigate.

Alternative library screening approaches for improved betaxanthin production

During the execution of the work presented here, results from a study by Savitskaya *et al.* (2019) was published using the same biosensor strategy to screen for improved betaxanthin production amongst mutant strain libraries. Their set-up was, however, different than the one employed here as they relied on traditional methods for yeast transformation. In a first round of screening in the Savitskaya *et al.* study, a pooled version of the deletion library was transformed with betaxanthin-synthesising genes and top producers were identified by fluorescent activated cell sorting (FACS). Although this avoided the issue of having to transform the ca. 4800 strains of the deletion library individually, only 85% coverage of the full library was possible. Furthermore, by using a pooled library the benefit of having a systematic array of strains is lost. With CRI-SPA, the information of each mutant is maintained and phenotypes of each strain can be scored, which also enable the use of systematic analyses such as the GO term enrichment analysis employed in our study.

The study by Savitskaya *et al.* (2019) further made use of multiple rounds of library screening to improve betaxanthin production. An iterative process like this would be highly interesting for future CRI-SPA based screens and we note that CRI-SPA is highly amenable to this type of set-up. With CRI-SPA, the possibility of CRISPR-Cas9 multiplexing means that genome-wide screens can be performed iteratively with the deletion library, in each cycle designing a new donor strain capable of not only transferring the biosynthetic pathway of interest, but also mutations identified as beneficial in prior rounds of screening. In the specific case of betaxanthin screening, it would be interesting to transfer the betaxanthin-

synthesising genes along with the *yor1Δ* mutation, as decreased export to the extracellular medium of the compound would facilitate simpler and more accurate assessment of pigment production in individual strains.

Evaluation of CRI-SPA in cell factory optimisation and future perspectives

This study set out to demonstrate the applicability of CRI-SPA in cell factory optimisation. In summary, CRI-SPA efficiently allowed us to simultaneously integrate the two genes for betaxanthin production in the strains of the yeast deletion collection in less than a week by using a simple replica pinning procedure. We analysed the betaxanthin production amongst mutants using a HT colorimetric assay and we validated our screen by confirming the phenotype of identified top-scoring mutations (of which one had already been identified by a previous study). Although further analysis of additional mutants and investigation of the biological effects of identified mutations is desirable, this initial study demonstrates the feasibility of using CRI-SPA for improving the performance of a cell factory.

Previously, as a result of endoreduplication, we observed a diploid recipient genome in the final strains generated by CRI-SPA (Chapter 2). Although Cas9 mediated gene transfer is efficient, there is always a risk that cells that underwent endoreduplication during haploidisation may end up heterozygous for the transferred genetic feature. This is important to keep in mind as a heterozygous and homozygous strain may have different phenotypes. If heterozygous cells are formed, their enrichment in the population of cells growing on the plates during the CRI-SPA procedure need to be counter-acted, because although initially constituting a minority, heterozygous cells may take over if possessing a growth advantage over homozygous cells. In our study, we supplemented growth media with elevated levels of tyrosine in order to lessen the growth disadvantage arising from synthesising betaxanthin, which is derived from intracellular tyrosine. A similar strategy may not always be applicable to other types of intended cell factory products, and heterologous production pathways tend to have a general negative effect on cellular fitness. We recommend that before conducting a CRI-SPA –based screen, one compares the competitive fitness of a homozygous diploid carrying the biochemical pathway of interest to a corresponding heterozygous diploid. Based on these observations, one may thereafter consider the use of an inducible system, where expression of pathway genes is only activated after completion of transfer. For a heterologous production pathway, a simple solution could be to use an inducible promoter of the gene responsible for catalysing the first step of the reaction pathway. Furthermore, in any HT study, biological replicates are key to increase the accuracy of analysis. We carried out our screen using four biological replicates and strains displaying large variations between replicates were excluded from analysis.

For future applications, our set-up with CRISPR-Cas9 offers exciting possibilities in terms of multiplexing. Metabolic engineering typically requires modification of several native host genes for efficient production when expressing heterologous biochemical pathways. In parallel to transferring genes for biochemical production, CRI-SPA allows simultaneous transfer of gene deletions and/or overexpressed alleles that have already been identified as key for improving production. Furthermore, as outlined above, multiplexing enables the yeast deletion collection to be used in multiple rounds of iterative screening, where beneficial mutations identified in prior screens are included in the transfer of subsequent rounds.

Utilisation of large strain libraries is not only limited by the availability of high-throughput methods for transformation and strain construction. A major bottle-neck in any cell factory optimisation process is the ability to screen and analyse a large number of strains and/or DNA constructs. In the set-up of this study, we made use of the fluorescent and colorimetric properties of betaxanthin to allow high-throughput screening of generated strains. Similar colorimetric assays would be possible for the production of other biotechnologically relevant compounds, including for example free fatty acids by implementation of a Nile Red staining protocol (66), and other colourful pigments such as indigo and anthocyanins derived from the aromatic amino acid pathway, and carotenoids derived from the mevalonate pathway. Furthermore, in order to meet the need for high-throughput screening approaches in modern metabolic engineering studies, much research is being devoted to the development of transcription-factor based biosensors, of which there are now several available for the use in yeast. These biosensor are able to convert intracellular concentrations of a target compound into a fluorescent signal and examples of these compounds include the aromatic molecules *cis,cis*-muconic acid and naringenin (67), the central precursor for fatty-acid derived compounds malonyl-CoA and fatty-acyl CoA (68–71), and isopentenyl diphosphate (IPP) (72), which is a key precursor in the production of isoprenoids. These biosensors could easily be transferred along with a production pathway or other genetic features by CRI-SPA to enable facile screening of large strain libraries in order to identify mutants with improved production capacities.

The study presented here demonstrates how the yeast deletion collection could be a more accessible tool in yeast cell factory optimisation. As discussed in Chapter 2, the method is also compatible with other genome-wide mutation libraries (73,74), which would be interesting to exploit for metabolic engineering purposes. Furthermore, yeast cell factory research is typically limited to the use of a small set of lab strains, thereby largely ignoring the great genetic diversity within the *S. cerevisiae* species. Here, we see extended possibilities for our method to tap into this resource by screening diverse strains backgrounds for their capacity of producing different compounds of industrial value. In contrast to competing methods for HT strain manipulation such as the one employed in Synthetic Genetic Array (SGA) analysis, CRI-SPA is independent of meiosis. In a scenario as the one outlined above, this is an advantage as it enables preservation of recipient genetic backgrounds that are non-isogenic to the UDS, whereas meiotic recombination would otherwise scramble the genome of the recipient and donor strain. Naturally, there are some restrictions in terms of recipient strain compatibility with the CRI-SPA method and in Chapter 4 of this thesis, we address these restrictions and suggest a possible solution. However, at this point, we note that the National collection of yeast cultures (NCYC) does for example harbour 28 haploid derivatives of *S. cerevisiae* isolates of diverse origin (75), that are fully compatible with the CRI-SPA method. These strains have a stable mating type and have further been rendered auxotrophic for uracil by insertion of the *kanMX* cassette at the *URA3* locus. The strains could easily be assessed for their capacity of producing various compounds by CRI-SPA mediated transfer of a set of biosynthetic genes and their diverse genetic backgrounds are likely to yield interesting variations in production performance. In this sense, a mating-based protocol could also offer additional advantageous as non-standard strain backgrounds may be difficult to transform in HT by the conventional lithium acetate procedure.

Methods

Strains and cultivation media

Strains used and constructed in this study are listed and described in Table 2. Strains were kept in YPD glycerol stocks (20 % (v/v)) at -80°C. Throughout the study, strains were grown in YP (10 g/L yeast extract, 20 g/L peptone) using either glucose (20 g/l), galactose (20 g/l) or raffinose (20 g/L) as carbon source or in synthetic complete (SC) medium (prepared as described by (76) but with 60 mg/L L-leucine). Galactose and raffinose solutions were sterilized by filtration. When growth on solid medium was required, 20 g/L agar was added. For selection when using antibiotic markers, plates were supplemented with G418 (200 mg/L), nourseothricin (NTC) (100 mg/L) and hygromycin (hyg) (200 mg/L) in different combinations. When grown in liquid, only half of the concentration of the respective antibiotic was used. For the CRI-SPA replica plating procedure, growth media was supplemented with 100 mg/L tyrosine. For *URA3* counter selection, SC plates were supplemented with 5-FOA (1 g/L) and uracil (30 mg/L). For cloning purposes, Escherichia coli was grown in LB medium (77) supplemented with 100 mg/L ampicillin and in the case for growth on solid medium, 20 g/L agar was added. The yeast genome-wide deletion collection was obtained from Invitrogen.

Table 2. Strains used and constructed in this study.

Strain name	Genotype	Source
BY4741	*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0	(78)
*Sc*_HO10	*MAT*α CEN$^{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2*	Chapter 2 of this thesis
*Sc*_HO20	*MAT*α CEN$^{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS)*	This study
*Sc*_HO25	*MAT*α CEN$^{1-16}$::pGal1-*KlURA3can1-100 his3-11,15 leu2-3,112 LYS2 met17 trp1-1 ura3-1 RAD5* X-3::p*TEF1*-SpCas9-t*CYC1*-loxP-*KlLEU2 KlURA3(dw XII-5 IS) XII-5(inner)Δ::tTDH3-BvCYP76AD1-pTPI1-pCCW12-MjDOD-tCYC1-loxP-NatMX-loxP*	This study
*Sc*_HO26	*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0 XII-5(inner)Δ::tTDH3-BvCYP76AD1-pTPI1-pCCW12-MjDOD-tCYC1-loxP-NatMX-loxP	This study
*Sc*_HO29	*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0 XII-5(inner)Δ::tTDH3-BvCYP76AD1-pTPI1-pCCW12-MjDOD-tCYC1-loxP-NatMX-loxP *cse2Δ::kanMX*	This study
*Sc*_HO30	*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0 XII-5(inner)Δ::tTDH3-BvCYP76AD1-pTPI1-pCCW12-MjDOD-tCYC1-loxP-NatMX-loxP *mlc2*Δ::kanMX	This study
*Sc*_HO31	*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0 XII-5(inner)Δ::tTDH3-BvCYP76AD1-pTPI1-pCCW12-MjDOD-tCYC1-loxP-NatMX-loxP *yor1*Δ::kanMX	This study

Design of gRNA sequences

gRNA target sequences in this study were generated using the Benchling CRISPR tool (https://benchling.com). Off-target effects were minimized by using the S288C reference genome and the algorithm from Doench *et al.* 2016 (79).

General cloning procedures

Primers used in this study are presented in Supplementary Table 4. All oligonucleotides were ordered from Integrated DNA Technologies (Leuven, Belgium). The pMEL12 plasmid (80) was obtained from EUROSCARF (Frankfurt, Germany). All cloning was done using uracil-specific excision reagent (USER™) enzyme from New England Biolabs and was carried out as previously described (81,82). DNA polymerases were purchased from Thermo Scientific and used according to the supplier's instructions. Standard PCR amplifications were run using Phusion Hot Start II High-Fidelity DNA Polymerase. Amplification of DNA for USER cloning was carried out using Phusion U Hot Start DNA Polymerase. In the case of colony PCR, Taq DNA Polymerase was used. Prior to cloning, amplified DNA was treated with DpnI (Thermo Scientific) and was column or gel purified using the illustra GFX PCR DNA and Gel Band Purification Kit (GE Lifesciences). All plasmids were verified by sequencing at Eurofins MWG Operon (Germany).

Plasmid construction

Plasmids used and constructed in this study are listed in Table 3. To construct a gRNA plasmid targeting an intergenic region downstream the XII-5 site(44), the target sequence in pCfB3050 (83) was first excluded by linearising the plasmid with uracil-containing primers HOP60 and HOP61. Then, a cassette containing a new 20 bp target sequence was generated by annealing the two single-stranded oligos HOP183 and HOP184 to form a double-stranded fragment with single-stranded 5' and 3' overhangs matching those of a USER-treated pCfB3050 backbone. The annealing was done by mixing the two oligos (100 µM) in equal volumes and then boiling for five minutes before allowing the mixture to cool to room temperature. The double-stranded cassette with the new target sequence was USER cloned into the linearised pCfB3050, yielding gRNA plasmid pHO25.

Table 3. Plasmids used and constructed in this study.

Plasmid name	Genetic element(s)	Type	Yeast marker	Source
pUG6	-	*E. coli* cloning vector	*kanMX*	(84)
pCfB390	-	Integrative (XI-3)	*Kl URA3*	(85)
pCfB2312	p*TEF1*-SpCas9(Hs opt.)-t*CYC1*	CEN/ARS	*kanMX*	(86)
pCfB3050	p*SNR52*-gRNA(XII-5)-t*SUP4*	2µ	*NatMX*	(83)
pMEL12	p*SNR52*-gRNA(*CAN1*)-t*SUP4*	2µ	*hphNT1*	(80)
Ant_E113	p*SNR52*-gRNA(XII-5)-tSUP4, *ARO7*^{G141S}-p*TEF1*-p*PGK1*-*ARO4*^{K229L}-t*TDH3*-*Bv*CYP76AD1-p*TPI1*-p*CCW12*-*Mj*DOD-t*CYC1*	2µ/integrative (XII-5)	*NatMX, URA3*	(45)
pHO25	p*SNR52*-gRNA(dw XII-5)-t*SUP4*	2µ	*NatMX, URA3*	This study
pHO29	p*SNR52*-gRNA(XII-5)-t*SUP4*	2µ	*hphNT1*	This study
pHO36	p*SNR52*-gRNA(XII-5)-t*SUP4*, t*TDH3*-*Bv*CYP76AD1-p*TPI1*-p*CCW12*-*Mj*DOD-t*CYC1*	2µ/integrative (XII-5)	*NatMX, URA3*	This study
pHO38	p*TEF1*-Cas9(Hs opt.)-t*CYC1*	CEN/ARS	*hphMX*	This study

To exchange the selective marker in the XII-5 targeting gRNA plasmid pCfB3050, the plasmid was linearised by uracil-containing primers HOP215 and HOP216, which excluded the *NatMX* cassette. The hphNT1 cassette was amplified

from pMEL12 with uracil-container HOP217 and HOP218. The *hphNT1* marker was then USER cloned into the linearised backbone to yield plasmid pHO29.

An integrative plasmid for inserting betaxanthin-synthesising genes in the XII-5 site was constructed using the Ant_E113 plasmid as template (45). PCR was run with primers HOP258 and HOP259 to exclude the *ARO4*^{K229L} and *ARO7*^{G141S} gene from the vector and the linearised plasmid, now only containing the *NatMX* marker and the CYP76AD1 and DOD gene, was closed by USER enzyme cloning, yielding plasmid pHO36.

The pHO38 Cas9 plasmid carrying the *hphNT1* marker was constructed by amplifying the Cas9 sequence of pCfB2312 (83) with primers HOP89 and HOP90 and the fragment was then USER cloned into the pSH69 plasmid from which the Cre gene had been excluded by linearisation with primers HOP284 and HOP285.

Strain construction

All yeast transformations were carried out using the LiAc / SS Carrier DNA / PEG method (87). When antibiotic resistance genes were used as markers, cells were recovered for 2 hours in YPD before being plated onto selective media. For all transformations, negative controls were included, where no DNA was added to the transformation mix. In each case between 200 -500 ng plasmid DNA was used and between 750 – 1500 ng of linear repair substrates. All linear DNA generated by PCR was gel purified prior to transformation. All strains was verified by colony PCR to ensure correct integration at the intended chromosomal locus.

To construct the new donor strain *Sc*_HO20, harbouring an additional *KlURA3* marker downstream the XII-5 integration site, strain *Sc*_HO10 was co-transformed with gRNA plasmid pHO25 and a linear *Kl URA3* repair substrate. The repair substrate was generated using plasmid pCfB390 (85) and primers HOP181 and HOP182 with 40 bp non-binding tails corresponding to the up- and downstream sequence of the targeted locus. The *Sc*_HO25 donor strain harbouring betaxanthin-synthesising genes in the XII-5 site, was then built by transforming *Sc*_HO20 with the integrative plasmid pHO36, which also harbours a gRNA expression cassette for targeting the XII-5 site.

A reference strain for production of betaxanthin was constructed by transforming strain BY4741 with Cas9 plasmid pHO38 and integrative plasmid pHO36 harbouring the betaxanthin genes and also a gRNA cassette for targeting the XII-5 site, resulting in strain *Sc*_HO26. Reverse engineered strains harbouring different gene deletions, were then constructed using the *Sc*_HO26 and *Sc*_HO27 strains. In each case, ca 500 bp of the sequence upstream and downstream of the targeted loci were amplified using S288C genomic DNA and the following primer pairs: HOP310/HOP311 and HOP312/HOP313 (*cse2*); HOP314/HOP315 and HOP316/HOP317 (*mlc2*); and HOP318/HOP319 and HOP320/HOP321 (*yor1*). For each primer pair, one of the primers contained a 40 bp non-binding tail corresponding to the 5' or 3' end of the kanMX cassette of pUG6 (84). This *kanMX* cassette was then amplified using primers HOP306 and HOP307, resulting in a fragment with 40 bp homology to the generated up- and downstream fragments. For each deletion, the up- and downstream sequence of the targeted locus were co-transformed with the *kanMX* fragment. For transformation of the betaxanthin reference strain *Sc*_HO26, this yielded strains *Sc*_HO29 (*cse2*Δ), *Sc*_HO30 (*mlc2*Δ) and *Sc*_HO31 (*yor1*Δ).

CRI-SPA high-throughput pin replication

Automatic pin replication was carried out using high-throughput pinning robot ROTOR HDA from Singer Instruments (United Kingdom), along with replica pinning pads (RePads) and rectangular petri dishes (PlusPlates) from the same company. To generate UDS lawns for mating to the strains of the deletion library, strain Sc_HO25 carrying gRNA plasmid pHO29 was inoculated in YPD supplemented with hygromycin and was grown overnight. For each plate of the deletion library, 700 uL of the UDS overnight culture was spread onto a rectangular YPD plate, onto which the deletion mutant strains were then replicated on top of to facilitate mating. The CRI-SPA replication procedure outlined in Figure 5B was then followed.

PhenoBooth colony screening and over-representation analysis of gene ontology terms

The CRI-SPA generated betaxanthin-producing library was analysed using the PhenoBooth colony counter from Singer Instruments. For each colony on the plates, an area was calculated the PhenoBooth image analysis software. The PhenoBooth software also gave a measure for colour intensity by calculating a redness value for each colony using the algorithm developed by Rodriguez *et al.* (88) The raw data was then processed using a Python script developed in-house, calculating average size and redness for the biological replicates of each strain present on the plates. Further, strains of which only one replicate had survived the process were omitted from analysis and a cut-off was applied in which strains with an average size smaller than within one standard deviation of the population median were excluded. Finally, the relative error for redness between replicates was calculated by dividing the standard deviation with the average and strains with a relative error larger than 10 % were omitted from analysis.

An over-representation analysis was then performed on the processed gene list, to determine if enriched (i.e. over-represented) or purified (i.e. under-represented) gene classifications were present amongst the genes. For classifying genes, gene ontologies (GO-terms) (89,90) for each *S. cerevisiae* gene were used, information which is stored in the gene database from the National Center for Biotechnology Information (NCBI) (91). As preliminary results did not yield significant results, group size was varied from 10 to 500 in increments of 10 to find the size that yielded significant results (p-value < 0.05), controlling the false-discovery rate using the Benjamini–Hochberg procedure (92). All computational analysis was conducted in python, using the mygene.info package (93,94) to convert gene ORFs to NCBI-compliant ids, and the GOATOOLS package (95) to perform all over/under-representation analysis. The Phenobooth output tables, together with all necessary scripts to reproduce the results presented here are available at https://github.com/computer-aided-biotech/phenobooth-analysis.

Analysis of fluorescence by flow cytometry

Analysed strains were cultivated in 5 mL YPD overnight. Cultures were diluted 1:100 in fresh YPD and were then grown for 48 h. For each strain, 1 mL was taken out and washed two times in MilliQ water before re-suspending in 500 uL PBS. 5 uL of this cell suspension was then added to 150 uL PBS in a microtiter plate and cells were analysed using the MACSQuant VYB flow cytometer (Miltenyi Biotec) equipped with a 488 nm laser (blue) and a 525-550 nm filter. For each strain, 10 000 single-cell events were recorded. Events were analysed with the FlowJo software (TreeStar Inc.) and a mean fluorescence intensity was calculated for the gated cell population of each sample. The presented data represents

the average of biological triplicates ($n = 3$) and the error bars represent one standard deviation (plus/minus). Strains were individually compared against the reference strain using a two-tailed, two-sample Student's t-test for samples with equal variance and differences were classified as significant for p-values < 0.05.

References

1. Nielsen J, Larsson C, van Maris A, Pronk J. Metabolic engineering of yeast for production of fuels and chemicals. Curr Opin Biotechnol. 2013;24(3):398–404.

2. Nielsen J, Keasling JD. Engineering Cellular Metabolism. Cell. 2016;164(6):1185–97.

3. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. Nature. 2013;496(7446):528–32.

4. Karim AS, Dudley QM, Jewett MC. Cell-Free Synthetic Systems for Metabolic Engineering and Biosynthetic Pathway Prototyping. In: Wittman C, Liao JC, editors. Industrial Biotechnology: Microorganisms. John Wiley & Sons; 2016. p. 125–48.

5. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science. 1999;285(5429):901–6.

6. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. Nature. 2002;418(6896):387–91.

7. Giaever G, Nislow C. The yeast deletion collection: A decade of functional genomics. Genetics. 2014;197(2):451–65.

8. Mira NP, Palma M, Guerreiro JF, Sá-Correia I. Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. Microb Cell Fact. 2010;9(1):79.

9. Endo A, Nakamura T, Ando A, Tokuyasu K, Shima J. Genome-wide screening of the genes required for tolerance to vanillin, which is a potential inhibitor of bioethanol fermentation, in *Saccharomyces cerevisiae*. Biotechnol Biofuels. 2008;1(1):3.

10. Fujita K, Matsuyama A, Kobayashi Y, Iwahashi H. The genome-wide screening of yeast deletion mutants to identify the genes required for tolerance to ethanol and other alcohols. FEMS Yeast Res. 2006;6(5):744–50.

11. Suzuki T, Yokoyama A, Tsuji T, Ikeshima E, Nakashima K, Ikushima S, *et al.* Identification and characterization of genes involved in glutathione production in yeast. J Biosci Bioeng. 2011;112(2):107–13.

12. Yoshida S, Yokoyama A. Identification and characterization of genes related to the production of organic acids in yeast. J Biosci Bioeng. 2012;113(5):556–61.

13. Kitagawa T, Kohda K, Tokuhiro K, Hoshida H, Akada R, Takahashi H, *et al.* Identification of genes that enhance cellulase protein production in yeast. J Biotechnol. 2011;151(2):194–203.

14. Hirasawa T, Takekuni M, Yoshikawa K, Ookubo A. Genome-wide identification of the targets for genetic manipulation to improve l- lactate production by *Saccharomyces cerevisiae* by using a single-gene deletion strain collection. J Biotechnol. 2013;168(2):185–93.

15. Özaydin B, Burd H, Lee TS, Keasling JD. Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. Metab Eng. 2013;15(1):174–83.

16. Cao M, Gao M, Suástegui M, Mei Y, Shao Z. Building microbial factories for the production of aromatic amino acid pathway derivatives : From commodity chemicals to plant-sourced natural products. Metab Eng. 2020;58(April 2019):94–132.

17. Rawat G, Tripathi P, Saxena RK. Expanding horizons of shikimic acid: Recent progresses in production and its endless frontiers in application and market trends. Appl Microbiol Biotechnol. 2013;97(10):4277–87.

18. Rinner U, Hudlicky T. Synthesis of Morphine Alkaloids and Derivatives. Top Curr Chem. 2012;(309):33–66.

19. Winter JM, Tang Y. Synthetic biological approaches to natural product biosynthesis. Curr Opin Biotechnol. 2012;23(5):736–43.

20. Chang MCY, Keasling JD. Production of isoprenoid pharmaceuticals by engineered microbes. Nat Chem Biol. 2006;2(12):674–81.

21. Jung ST, Lauchli R, Arnold FH. Cytochrome P450: Taming a wild type enzyme. Curr Opin Biotechnol. 2011;22(6):809–17.

22. Curran KA, Leavitt JM, Karim AS, Alper HS. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab Eng. 2013;15(1):55–66.

23. Niu W, Draths KM, Frost JW. Benzene-free synthesis of adipic acid. Biotechnol Prog. 2002;18(2):201–11.

24. Lu R, Lu F, Chen J, Yu W, Huang Q, Zhang J, *et al.* Production of Diethyl Terephthalate from Biomass-Derived Muconic Acid. Angew Chemie - Int Ed. 2016;55(1):249–53.

25. Hansen EH, Møller BL, Kock GR, Bünner CM, Kristensen C, Jensen OR, *et al.* De novo biosynthesis of Vanillin in fission yeast (schizosaccharomyces pombe) and baker's yeast (*Saccharomyces cerevisiae*). Appl Environ Microbiol. 2009;75(9):2765–74.

26. Priefert H, Rabenhorst J, Steinbüchel A. Biotechnological production of vanillin. Appl Microbiol Biotechnol. 2001;56(3–4):296–314.

27. Suástegui M, Yu Ng C, Chowdhury A, Sun W, Cao M, House E, *et al.* Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in *Saccharomyces cerevisiae* for high production of polymer and drug precursors. Metab Eng. 2017;42(December 2016):134–44.

28. Adelfo Escalante A, Carmona SB, Diaz Quiroz DC, Bolivar F. Current perspectives on applications of shikimic and aminoshikimic acids in pharmaceutical chemistry. Res Reports Med Chem. 2014;35.

29. Germann SM, Baallal Jacobsen SA, Schneider K, Harrison SJ, Jensen NB, Chen X, *et al.* Glucose-based microbial production of the hormone melatonin in yeast *Saccharomyces cerevisiae*. Biotechnol J. 2016;11(5):717–24.

30. Zhang J, Wu C, Sheng J, Feng X. Molecular basis of 5-hydroxytryptophan synthesis in: *Saccharomyces cerevisiae*. Mol Biosyst. 2016;12(5):1432–5.

31. Kim B, Cho BR, Hahn JS. Metabolic engineering of *Saccharomyces cerevisiae* for the production of 2-phenylethanol via Ehrlich pathway. Biotechnol Bioeng. 2014;111(1):115–24.

32. Koopman F, Beekwilder J, Crimi B, Houwelingen A Van, Hall RD, Bosch D, *et al.* De novo production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. Microb Cell Fact. 2012;11(155):1–15.

33. Rodriguez A, Strucko T, Gustav S, Kristensen M, Killerup D, Forster J, *et al.* Metabolic engineering of yeast for fermentative production of flavonoids. Bioresour Technol. 2017;245:1645–54.

34. Li M, Kildegaard KR, Chen Y, Rodriguez A, Borodina I, Nielsen J. De novo production of resveratrol from glucose or ethanol by engineered *Saccharomyces cerevisiae*. Metab Eng. 2015;32:1–11.

35. Fossati E, Ekins A, Narcross L, Zhu Y, Falgueyret JP, Beaudoin GAW, *et al.* Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in *Saccharomyces cerevisiae*. Nat Commun. 2014;5:3283.

36. Li Y, Li S, Thodey K, Trenchard I, Cravens A, Smolke CD. Complete biosynthesis of noscapine and halogenated alkaloids in yeast. Proc Natl Acad Sci U S A. 2018;115(17):E3922–31.

37. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD. Complete biosynthesis of opioids in yeast. Science (80-). 2015;349(6252):1095–100.

38. Huccetogullari D, Luo ZW, Lee SY. Metabolic engineering of microorganisms for production of aromatic compounds. Microb Cell Fact. 2019;18:41.

39. Bender D. Amino Acid Metabolism. 3rd ed. Chichester, West Sussex: Wiley- Blackwell; 2012.

40. Braus GH. Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway. Microbiol Rev. 1991;55(3):349–70.

41. DeLoache WC, Russ ZN, Narcross L, Gonzales AM, Martin VJJ, Dueber JE. An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nat Chem Biol. 2015;11:465–71.

42. Salat D, Tolosa E. Levodopa in the treatment of Parkinson's disease: Current status and new developments. J Parkinsons Dis. 2013;3(3):255–69.

43. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013;31(9):827–32.

44. Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH, *et al.* Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metab Eng. 2012;14(2):104–11.

45. Coumou H. Engineering baker's yeast for the production of aromatic plant-derived compounds. Technical University of Denmark; 2016.

46. Decottignies A, Grant AM, Nichols JW, De Wet H, McIntosh DB, Goffeau A. ATPase and multidrug transport activities of the overexpressed yeast ABC protein Yor1p. J Biol Chem. 1998;273(20):12612–22.

47. Savitskaya J, Protzko RJ, Li FZ, Arkin AP, Dueber JE. Iterative screening methodology enables isolation of strains with improved properties for a FACS-based screen and increased L-DOPA production. Sci Rep. 2019;9(1):1–10.

48. Kang JS, Kim SH, Hwang MS, Han SJ, Lee YC, Kim YJ. The Structural and Functional Organization of the Yeast Mediator Complex. J Biol Chem. 2001;276(45):42003–10.

49. Malik S, Roeder RG. The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. Nat Rev Genet. 2010;11(11):761–72.

50. Luo J, Vallen EA, Dravis C, Tcheperegine SE, Drees B, Bi E. Identification and functional analysis of the essential and regulatory light chains of the only type II myosin Myo1p in *Saccharomyces cerevisiae*. J Cell Biol. 2004;165(6):843–55.

51. Hamey JJ, Wilkins MR. Methylation of Elongation Factor 1A: Where, Who, and Why? Trends Biochem Sci. 2018;43(3):211–23.

52. Han SJ, Lee JS, Kang JS, Kim YJ. Med9/Cse2 and Gal11 Modules are Required for Transcriptional Repression of Distinct Group of Genes. J Biol Chem. 2001;276(40):37020–6.

53. Zhang F, Sumibcay L, Hinnebusch AG, Swanson MJ. A Triad of Subunits from the Gal11/Tail Domain of Srb Mediator Is an In Vivo Target of Transcriptional Activator Gcn4p. Mol Cell Biol. 2004;24(15):6871–86.

54. Poss ZC, Ebmeier CC, Taatjes DJ. The Mediator complex and transcription regulation. Crit Rev Biochem Mol Biol. 2013;48(6):575–608.

55. Xiao Z, McGrew JT, Schroeder AJ, Fitzgerald-Hayes M. CSE1 and CSE2, two new genes required for accurate mitotic chromosome segregation in *Saccharomyces cerevisiae*. Mol Cell Biol. 1993;13(8):4691–702.

56. Xiao ZX, Fitzgerald-Hayes M. Functional interaction between the CSE2 gene product and centromeres in *Saccharomyces cerevisiae*. J Mol Biol. 1995;248(2):255–63.

57. Lanouette S, Mongeon V, Figeys D, Couture JF. The functional diversity of protein lysine methylation. Mol Syst Biol. 2014;10(4):1–26.

58. Cavallius J, Popkie AP, Merrick WC. Site-directed mutants of post-translationally modified sites of yeast eEF1A using a shuttle vector containing a chromogenic switch. Biochim Biophys Acta - Gene Struct Expr. 1997;

59. Martín-Granados C, Riechers SP, Stahl U, Lang C. Absence of see1p, a widely conserved *Saccharomyces cerevisiae* protein, confers both deficient heterologous protein production and endocytosis. Yeast. 2008;

60. Lipson RS, Webb KJ, Clarke SG. Two novel methyltransferases acting upon eukaryotic elongation factor 1A in *Saccharomyces cerevisiae*. Arch Biochem Biophys. 2010;

61. Li Z, Gonzalez PA, Sasvari Z, Kinzy TG, Nagy PD. Methylation of translation elongation factor 1A by the METTL10-like See1 methyltransferase facilitates tombusvirus replication in yeast and plants. Virology. 2014;

62. Kushner DB, Lindenbach BD, Grdzelishvili VZ, Noueiry AO, Paul SM, Ahlquist P. Systematic, genome-wide identification of host genes affecting replication of a positive-strand RNA virus. Proc Natl Acad Sci U S A. 2003;100(26):15764–9.

63. Anderson RM, Bitterman KJ, Wood JG, Medvedik O, Sinclair DA. Nicatinamide and PNC1 govern lifespan extension by calorie restriction in *Saccharomyces cerevisiae*. Nature. 2003;

64. White JT, Cato T, Deramchi N, Gabunilas J, Roy KR, Wang C, *et al.* Protein Methylation and Translation: Role of Lysine Modification on the Function of Yeast Elongation Factor 1A. Biochemistry. 2019;58(49):4997–5010.

65. Kubota H, Obata T, Ota K, Sasaki T, Ito T. Rapamycin-induced translational derepression of GCN4 mRNA involves a novel mechanism for activation of the eIF2α kinase GCN2. J Biol Chem. 2003;278(23):20457–60.

66. Shi S, Ji H, Siewers V, Nielsen J. Improved production of fatty acids by *Saccharomyces cerevisiae* through screening a cDNA library from the oleaginous yeast Yarrowia lipolytica. FEMS Yeast Res. 2015;16(1):1–10.

67. Skjoedt ML, Snoek T, Kildegaard KR, Arsovska D, Eichenberger M, Goedecke TJ, *et al.* Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. Nat Chem Biol. 2016;12(11):951–8.

68. Xu P, Li L, Zhang F, Stephanopoulos G, Koffas M. Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. Proc Natl Acad Sci U S A. 2014;

69. David F, Nielsen J, Siewers V. Flux Control at the Malonyl-CoA Node through Hierarchical Dynamic Pathway Regulation in *Saccharomyces cerevisiae*. ACS Synth Biol. 2016;5(3):224–33.

70. Zhang F, Carothers JM, Keasling JD. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat Biotechnol. 2012;30(4):354–9.

71. Teo WS, Hee KS, Chang MW. Bacterial FadR and synthetic promoters function as modular fatty acid sensor-regulators in *Saccharomyces cerevisiae*. Eng Life Sci. 2013;13(5):456–63.

72. Chou HH, Keasling JD. Programming adaptive control to evolve increased metabolite production. Nat Commun. 2013;

73. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat Methods. 2008;5(8):711–8.

74. Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, *et al.* Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell. 2006;21(3):319–30.

75. Cubillos FA, Louis EJ, Liti G. Generation of a large set of genetically tractable haploid and diploid Saccharomyces strains. FEMS Yeast Res. 2009;9(8):1217–25.

76. Sherman F, Fink G., Hicks J. Laboratory Course Manual for Methods in Yeast Genetics. Cold Spring Harb Lab. 1986;

77. BERTANI G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. J Bacteriol. 1951;62(3):293–300.

78. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast. 1998;14(2):115–32.

79. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol. 2016;34(2):184–91.

80. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NGA, van den Broek M, *et al.* CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res. 2015;15(2):1–15.

81. Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA. USER fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res. 2007;35(7):0–5.

82. Nour-Eldin HH, Hansen BG, Nørholm MHH, Jensen JK, Halkier BA. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res. 2006;34(18).

83. Jessop-Fabre MM, Jakočiūnas T, Stovicek V, Dai Z, Jensen MK, Keasling JD, *et al.* EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. Biotechnol J. 2016;11(8):1110–7.

84. Güldener U, Heck S, Fiedler T, Beinhauer J, Hegemann JH. A new efficient gene disruption cassette for repeateduse in budding yeast. Nucleic Acids Res. 1996;24(13):2519–24.

85. Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, *et al.* EasyClone: Method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. FEMS Yeast Res. 2014;14(2):238–48.

86. Stovicek V, Borodina I, Forster J. CRISPR-Cas system enables fast and simple genome editing of industrial *Saccharomyces cerevisiae* strains. Metab Eng Commun. 2015;2:13–22.

87. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc / SS carrier DNA / PEG method. Nat Protoc. 2008;2(1):31–5.

88. Rodriguez JD, Johnston PR, Ousler GW, Smith LM, Abelson MB. Automated grading system for evaluation of ocular redness associated with dry eye. Clin Ophthalmol. 2013;7:1197–204.

89. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, *et al.* Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25–9.

90. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, *et al.* The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019;47(D1):D330–8.

91. Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, *et al.* Gene: A gene-centered information resource at NCBI. Nucleic Acids Res. 2015;43(D1):D36–42.

92. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995;57(1):289–300.

93. Wu C, MacLeod I, Su AI. BioGPS and MyGene.info: Organizing online, gene-centric information. Nucleic Acids Res. 2013;41(1):561–5.

94. Xin J, Mark A, Afrasiabi C, Tsueng G, Juchler M, Gopal N, *et al.* High-performance web services for querying gene and variant annotation. Genome Biol. 2016;17(1):1–7.

95. Klopfenstein D V., Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, *et al.* GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 2018;8(1):1–17.

Supplementary information to Chapter 3

$Plate_{384}$ 8

Plate 29	Plate 30	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	
Plate 31	Plate 32	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	12	12	
A	A	MRPL22	SWM1	YNL176C	EXG2	NOP13	YDR262W	MDG1	DIN7	YNL170W	AKR1	YNL171C	PEX10	PSD1	YDR266C	FMP41	YDR269C	SKO1	CCC2	BNI5	YDR271C	YNL165W	GLO2	IBD2	DON1	A
A	A	SEC28	YOR1	YIL077C	BGL2	AIR1	YGR283C	SDS3	ERV29	YIL086C	ZUO1	AIM19	BIO2	AVT7	YGR287C	ICE2	MAL13	RSM25	YGR290W	PRK1	ECM12	YIL096C	YHR039C-B	FYV10	YHR079C-B	B
B	B	RPL42A	YDR274C	YGP1	BSC2	ASI2	PMP3	YNL157W	MTH1	NSG2	YDR278C	YNL155W	RNH202	YCK2	PHM6	ALF1	YDR282C	DPH2	DPP1	PEX1	ZIP1	PTK1	YDR286C	YKL199C	INM2	C
B	B	YIA6	EST3	NAS2	MET28	URM1	YAP5	FAA3	MUC1	DOT5	YIR020C	YIL015C-A	YIR020W-B	NOT3	MRS1	PKP1	YIR024C	SYG1	MND2	RPL34B	YVH1	YIL054W	DAL1	YIL055C	DAL4	D
C	C	YKL200C	RTT103	LOS1	YDR290W	ADD66	HRQ1	AIM27	SSD1	0	DPL1	CBT1	0	TRP3	HDA2	SAC1	SUR2	DOA1	ATP5	YRA2	CPR5	URA1	HNT2	JEN1	YDR306C	E
C	C	YIL059C	DAL2	YIL060W	DAL7	YIL067C	MGA2	RPS24B	LYS1	MAM33	YIR035C	PCI8	IRC24	0	HYR1	SER33	0	YIL089W	GTT1	YIL092W	YPS6	YFL006W	YIR042C	HXT10	YKL033W-A	F
D	D	SRY1	YDR307W	MCH2	GIC2	YKL222C	SUM1	VPS1	SSF2	OSH6	PIB1	YKR005C	RAD34	MEH1	IPK1	FOX2	OMS1	YKR011C	HIM1	YKR012C	MCM21	PRY2	YDR319C	YPT52	SWA2	G
D	D	YFL015C	YKL162C-A	PAU5	DID2	GAT1	CCP1	BUD27	GPT2	BST1	MET1	STE2	YKR070W	GYP8	SIS2	CAF16	YKR073C	AGX1	AIM29	HAC1	YKR075C	YFL032W	ECM4	YFL034W	MSA2	H
E	E	YKR015C	ASP1	FCJ1	MRPL35	YKR017C	PEP7	YKR018C	PEX3	VPS51	UBX5	ALY1	IRC3	DBP7	YDR333C	GCN3	SWR1	GMH1	MSN5	SPO14	YDR336W	YKR032W	MRPS28	YKR033C	YIL001W	I
E	E	YFL035C-B	YKR078W	RPO41	MTD1	YFL040W	NUP133	FET5	HBS1	YFL043C	ARG7	OTU1	RIM9	FMP32	KAR5	RGD2	UBX4	EMP47	AVO2	SWP82	NAT4	ALR2	MOT3	YFL051C	TVP18	J
F	F	OPI8	INP51	UTH1	EPS1	YKR043C	TIR3	UIP5	YIL012W	YKR045C	PDR11	YKR047W	MNT3	NAP1	BAR1	FMP46	SNL1	TRK2	VID28	YKR051W	HIS6	MRS4	YKE4	DYN1	YIL024C	K
F	F	YFL052W	ABF2	DAK2	IRC21	YFL054C	YMR075C-A	AGP3	RCO1	AAD6	VPS20	LOC1	CTF18	YFR006W	NAM7	YFH7	ISF1	FAR7	YMR082C	GCN20	ADH3	UBP6	YMR084W	YFR012W	YMR085W	L
G	G	RHO4	YIL025C	TRM2	KRE27	RPS21A	YIL028W	GLG1	YIL029C	TIF1	YIL032C	UTP30	CAP2	KTR2	CKA1	OAF3	CST6	PAM17	PRM2	PEX5	TED1	VHS1	APQ12	YDR248C	GVP36	M
G	G	CMK1	YMR086C-A	GSY1	YMR086W	YFR016C	YMR087W	YFR017C	VBA1	YFR018C	YTA12	YFR020W	YMR090W	ATG18	NPL6	ROG3	AIP1	PES4	PDR18	LSB3	YNR071C	ULI1	HXT17	RPL2A	YNR073C	N
H	H	YDR249C	CBR1	0	0	YDR250C	AGE2	PAM1	PIG2	BTT1	DFG10	MET32	PCL7	CHL4	RHR2	RMD5	YIL057C	CTA1	SEE1	RKM4	FIS1	HSP78	HOP1	YAP6	SPO22	O
H	H	GND2	AIF1	0	0	YGR259C	COS10	TNA1	YOL013W-A	APL6	ADH1	SAY1	YOL087C	YGR266W	MPD2	HUA1	HAL9	YGR269W	MSH2	YTA7	SPO21	RTT102	YOL092W	SCW4	TRM10	P
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	

Supplementary Figure 2. Positions of mutant strains on Plate 8 of the yeast deletion library when in a 384 density format.

Supplementary Table 1. Top 150 mutations for colour intensity of betaxanthin-producing strains.

Top 150 (highest to lowest)					
Rank	**Gene name**	**Rank**	**Gene name**	**Rank**	**Gene name**
1	CSE2	51	FMS1	101	KEL1
2	MLC2	52	HCH1	102	DYN1
3	YOR1	53	YJL175W	103	SSU1
4	NTA1	54	TPK1	104	YBR032W
5	YOL153C	55	FMC1	105	END3
6	CRG1	56	ASE1	106	MRPS35
7	NCS2	57	TMA108	107	MKK1
8	ATX1	58	BNI5	108	ARR3
9	ETP1	59	NOT3	109	RTR1
10	YLR446W	60	SAW1	110	FMP46
11	SSM4	61	THI21	111	YEL047C
12	JSN1	62	YEL028W	112	OST5
13	UNG1	63	SPL2	113	NNT1
14	POA1	64	YNL010W	114	SKY1
15	GAS5	65	DAL4	115	PUS1
16	YDR370C	66	YNL115C	116	PDR18
17	YMR279C	67	PNC1	117	YHK8
18	HDA3	68	YRO2	118	YBR271W
19	NPR3	69	RSM27	119	YDC1
20	YNL165W	70	EPT1	120	AGX1
21	FMP41	71	YPT31	121	COQ6
22	NUC1	72	TIP1	122	PNS1
23	SPO22	73	YML102C-A	123	PGD1
24	PPE1	74	IOC3	124	YGR017W
25	YML048W-A	75	MLF3	125	MPH1
26	TGL2	76	NRG2	126	ARN2
27	FIG1	77	YIP3	127	FRT2
28	SET2	78	YNL024C	128	YNL040W
29	ECI1	79	YIL015C-A	129	RPN4
30	YER010C	80	DAL80	130	SAF1
31	PKP1	81	ALE1	131	VID30
32	SKO1	82	AVT3	132	YPL264C
33	EMP70	83	YHL037C	133	PAC2
34	PRM2	84	YGR050C	134	YMC1
35	YPR174C	85	EXG2	135	YDL186W
36	SEE1	86	MIP6	136	YGR093W
37	NDT80	87	ASN1	137	HXT10
38	PAP2	88	TPM2	138	SLM2
39	GVP36	89	YKL202W	139	YJL163C
40	MUP3	90	EDS1	140	YBR139W
41	RRT13	91	RPS10B	141	YNL095C
42	RAD16	92	SGN1	142	SPO14
43	YPL088W	93	BNA2	143	ALD2
44	TIR1	94	YMR114C	144	FLD1
45	UBA4	95	SIM1	145	TSR3
46	CPR8	96	UTR1	146	PET100
47	YNL190W	97	RLM1	147	YMR153C-A
48	YJL181W	98	KSS1	148	YHL012W
49	YMR294W-A	99	IBD2	149	SPO73
50	TAE1	100	ROM1	150	RKM1

Supplementary Table 2. Bottom 150 mutations for colour intensity of betaxanthin-producing strains.

Bottom 150 (lowest to highest)				
Rank	**Gene name**	**Rank**	**Gene name**	**Rank**
1	IBA57	51	GID7	101
2	OPY1	52	RTT10	102
3	YCR090C	53	YDL133W	103
4	HTD2	54	YDL094C	104
5	DIA3	55	DIG2	105
6	MRPS9	56	THI22	106
7	TAF14	57	YGR039W	107
8	YMR052C-A	58	VTS1	108
9	RIM8	59	GLC3	109
10	THI72	60	GRX2	110
11	ECM11	61	YBR284W	111
12	RPB9	62	IRC2	112
13	PCS60	63	AAT1	113
14	YDR509W	64	MSC3	114
15	CTF19	65	YBR300C	115
16	NRG1	66	YEF1	116
17	ARO2	67	YGR031W	117
18	STB2	68	YDL121C	118
19	SNF7	69	YML096W	119
20	VPS3	70	RSC1	120
21	KIN82	71	BAT2	121
22	KAR4	72	FLO10	122
23	UBA3	73	YJL206C	123
24	SSH4	74	AIM14	124
25	BFA1	75	HIS2	125
26	FAR3	76	NAT3	126
27	MPC54	77	YFL052W	127
28	NGR1	78	CYB5	128
29	CSM3	79	EUG1	129
30	ROM2	80	YOR331C	130
31	YLR173W	81	YMR118C	131
32	PKH3	82	PGM2	132
33	DUN1	83	VAC14	133
34	UPF3	84	HLR1	134
35	ALD3	85	YOL159C	135
36	YSY6	86	YFL063W	136
37	RPL22B	87	YSW1	137
38	APL3	88	YBR028C	138
39	YCL001W-A	89	YKU80	139
40	AFR1	90	SPG4	140
41	RSM28	91	YRB30	141
42	NUT1	92	RTC4	142
43	PHO8	93	HFA1	143
44	MTC5	94	FKH1	144
45	YDL011C	95	YCL002C	145
46	STP1	96	YCL006C	146
47	YDL057W	97	YER038W-A	147
48	YNR061C	98	HOR2	148
49	YHR035W	99	SVP26	149
50	YPK2	100	MET10	150

Supplementary Table 3. Genes out of the top and bottom 150 associated with the GO term "metabolic process".

Gene name	Gene list	Rank
UBA4	Top 150	45
EPT1	Top 150	70
EXG2	Top 150	85
SIM1	Top 150	95
HFA1	Bottom 150	93

Supplementary Table 4. **Primers used in this study.** USER cloning tails are indicated in bold; gRNA protospacer sequences are underlined; long non-binding tails are written in lowercase letters.

Primer name	Sequence (5' to 3')	Purpose
Primers for cloning		
HOP60	**GTTTTAGAGCU**AGAAATAGCAAG	Linearizing plasmid pCfB3050 for cloning of new gRNA target sequence
HOP61	**GATCATTTAU**CTTTCACTGC	Linearizing plasmid pCfB3050 for cloning of new gRNA target sequence
HOP183	ATTCATGTGATAATTTGACG**GTTTTAGAGCT**	Cloning gRNA target sequence for downstream XII-5, (+) strand
HOP184	CGTCAAATTATCACATGAAT**GATCATTTAT**	Cloning gRNA target sequence for downstream XII-5, (-) strand
HOP215	**ATCGCACGU**GATTCGATACTAACGCCGC	Construction of plasmid pHO29
HOP216	**ATCGCGTGU**GACGAGGCAAGCTAAACAG	Construction of plasmid pHO29
HOP217	**ACGTGCGAU**TCTCGAGAGCTCGTTAAAGC	Construction of plasmid pHO29
HOP218	**ACACGCGAU**AGAGATCTGTTTAGCTTGCCTCG	Construction of plasmid pHO29
HOP258	**ATTCTGCAGGU**AGGGAAAGATATGAG	Construction of plasmid pHO36
HOP259	**ACCTGCAGAAU**CGAGCTCGCTGAGGAC	Construction of plasmid pHO36
HOP89	**ACACGCGAU**GCTGGAGCTCATAGCTTCA	Construction of plasmid pHO38
HOP90	**ACGTGCGAU**CTATAGGGCGAATTGGGTAC	Construction of plasmid pHO38
HOP284	**ATCGCACGU**CGCAGGAAAGAACATGTGAGC	Construction of plasmid pHO38
HOP285	ATCGCGTGUGATTAAGGGTTCTCGAGAGCTC	Construction of plasmid pHO38
HOP296	**AGTCTGTGCAU**CGAAGTCTTGTTTAAAGATTACGG	Construction of plasmid pHO50
HOP297	**ATGCACAGACU**TATGCTATACGAAGTTATTAGGTCTAG	Construction of plasmid pHO50
Primers for sequencing		
EDR441	AACCGTATTACCGCCTTTGAGTGAGCT	Sequencing gRNA plasmids
HOP50	GACTCGAGATTATTCCTTTGC	Sequencing gRNA plasmids
HOP51	CAGTTTGCCAGTGATACAC	Sequencing gRNA plasmids
HOP52	GAC AAT TCA ACG CGT CTG	Sequencing gRNA plasmids and pHO50
HOP197	ATCGCACGUCACTGGATGGCGGCGTTAG	Sequencing gRNA plasmids
HOP14	GCATGATTTGATGGCTGTACC	Sequencing pHO36
HOP15	CCTTTTCCATCATTTCAGG	Sequencing pHO36
HOP16	GGTAAAGGTTTAGGACCTGG	Sequencing pHO36

Supplementary Table 4 (cont.)

Primer name	Sequence (5' to 3')	Purpose
HOP17	CGAATCGGACGACGAATCG	Sequencing pHO36
HOP48	GCGTGTACGCATGTAAC	Sequencing pHO36
HOP52	GAC AAT TCA ACG CGT CTG	Sequencing pHO36
HOP88	GGTGTCGGTGGTGAAGGAC	Sequencing pHO36
HOP258	ATTCTGCAGGUAGGGAAAGATATGAG	Sequencing pHO36
HOP259	ACCTGCAGAAUCGAGCTCGCTGAGGAC	Sequencing pHO36
Ant_P516	CGCGTCAGCTGAAGCTTCGTAC	Sequencing pHO36
Ant_P517	AAACGAGCTCTCGAGAACCC	Sequencing pHO36
PR_DIV105	CACGCGAUCTTCGAGCGTCCCAAAACC	Sequencing pHO36
HV_20	ATGATTAAATGCTACTTCAACAGGAATGGACCAGCTTCAGAACCT	Sequencing pHO36
HOP89	ACACGCGAUGCTGGAGCTCATAGCTTCA	Sequencing pHO38
HOP90	ACGTGCGAUCTATAGGGCGAATTGGGTAC	Sequencing pHO38
HOP93	AATTCAAAGTTCTGGGCAATACCG	Sequencing pHO38
HOP94	TGGCCGAAGATGCCAAGCTTCAAC	Sequencing pHO38
HOP95	CAGTCCTTCATCGAAAGGATGAC	Sequencing pHO38
HOP96	TGCAGACCGTTAAGGTCGTGGATG	Sequencing pHO38
HOP97	ATTACTCTGAAGTCTAAGCTGGTC	Sequencing pHO38
HOP98	AAAGACCTCATCATTAAGCTTCCC	Sequencing pHO38
HOP99	TGCTGTGGCGATCGGTATTG	Sequencing pHO38
Primers for strain construction		
HOP181	gtatacctttcatttacaaaatgtcaaagactttacacagGGGTTCTCGAGAGCTCG	Amplification of *Kl URA3* from pCfB390
HOP182	cgcaatgacatagcttagagcagtcagaaaagtgattgtaGGTCTAGAGATCCCAATACA	Amplification of *Kl URA3* from pCfB390
HOP306	GTATGCTATACGAAGTTATTAGGTCTAGAG	Amplification of kanMX from pUG6
HOP307	GTTCTCGAGAGCTCGTTTTCG	Amplification of kanMX from pUG6
HOP310	GGTGTCAGCGAAAGCTCG	Amplification of *CSE2* upstream sequence
HOP311	ctaaacagatctctagacctaataacttcgtatagcatacCGACTTTCAGTGCTCGTGTC	Amplification of *CSE2* upstream sequence
HOP312	taacgccgccatccagtgtcgaaaacgagctctcgagaacGCATACCACATCTTCAAACG	Amplification of *CSE2* downstream sequence
HOP313	CAGGATGAAGATAGGCAATATAGC	Amplification of *CSE2* downstream sequence
HOP314	CGAAAGTATGTCTGGGACAACC	Amplification of *MLC2* upstream sequence
HOP315	ctaaacagatctctagacctaataacttcgtatagcatacCTACTTCAAGTTTGCTCAATAAACG	Amplification of *MLC2* upstream sequence
HOP316	taacgccgccatccagtgtcgaaaacgagctctcgagaacCGTGGTTCTCTCGTCCAG	Amplification of *MLC2* downstream sequence
HOP317	GCATGAGCAAATAAGAAGAAAGAC	Amplification of *MLC2* downstream sequence
HOP318	GCTTCTATTTTCCTCCCCACC	Amplification of *YOR1* upstream sequence
HOP319	ctaaacagatctctagacctaataacttcgtatagcatacCCTCGTTCGTGACACTACTC	Amplification of *YOR1* upstream sequence
HOP320	taacgccgccatccagtgtcgaaaacgagctctcgagaacCGATGGTACAAATTAGTACTAGAAAAG	Amplification of *YOR1* downstream sequence
HOP321	CCTTCTGTTTAGGTGAGGAATAAGC	Amplification of *YOR1* downstream sequence

Supplementary Table 4 (cont.)

Primer name	Sequence (5' to 3')	Purpose
Primers for strain verification		
HOP185	CAGTGTCATATAGAAAGACAGCGCTC	Verifying integration of Kl URA3 dw XII-5
HOP186	GCTAAGAGCGAAGTGGGAGG	Verifying integration of Kl URA3 dw XII-5
HOP187	GTTTCGATGCAACCGGACTTGC	Verifying integration of Kl URA3 dw XII-5
HOP196	ACGATATGGGAGGAAGAGAAGAAGG	Verifying integration of Kl URA3 dw XII-5
HOP271	CCAGGATAACGATTCTGAAGTTAC	Verifying integrations at XII-5
Ant_P497	CAGTCAGCTCTTTCGTAGAC	Verifying integrations at XII-5
Ant_P498	GAACAAATGAAGATAATGGAGC	Verifying integrations at XII-5
Ant_P517	AAACGAGCTCTCGAGAACCC	Verifying integrations at XII-5
PR_DIV077	ATCAGTACUGACAATAAAAAGATTCTTGTT	Verifying integrations at XII-5
PR_DIV682	ATTGTGUTTTGATAGTTGTTCAATTGATTGAA	Verifying integrations at XII-5
HOP308	CGAAACGTGAGTCTTTTCCTTACC	Verifying deletion of CSE2, MLC2 and YOR1, binds to kanMX
HOP309	GCAGTTTCATTTGATGCTCGATGAG	Verifying deletion of CSE2, MLC2 and YOR1, binds to kanMX
HOP370	CACGATAATACATCTTCCTTTACATCTTTCC	Verifying deletion of *CSE2*, binds upstream the ORF
HOP371	GTTGGATTAGTTGAAAAGAACGACCAAGC	Verifying deletion of *CSE2*, binds downstream the ORF
HOP372	GATTCGTAGGCAGCAGAATCTGG	Verifying deletion of *CSE2*, binds in the middle of the ORF, reverse direction
HOP373	GGATGATCTTTATCGCAAGTTACAACG	Verifying deletion of *CSE2*, binds in the middle of the ORF, forward direction
HOP374	CTGCATATTGGCCACACAGG	Verifying deletion of *MLC2*, binds upstream the ORF
HOP375	CCATCGTTACTGGTGGTAATGG	Verifying deletion of *MLC2,* binds downstream the ORF
HOP376	CGTTAGCGATTCACTATGGTCC	Verifying deletion of *MLC2*, binds in the middle of the ORF, reverse direction
HOP377	GACCGAAGAGAGAACTTTCAGAGG	Verifying deletion of *MLC2*, binds in the middle of the ORF, forward direction
HOP378	CACGGTTCTCTACTAAATGCTTGC	Verifying deletion of *YOR1*, binds upstream the ORF
HOP379	GGACTGTGACTTTTCATGAACGC	Verifying deletion of *YOR1,* binds downstream the ORF
HOP380	GGTAATCGTCATATTCGTATATAGCAACG	Verifying deletion of *YOR1*, binds in the middle of the ORF, reverse direction
HOP381	CATGTGTTCTAGATCTGGTATTGTGG	Verifying deletion of *YOR1*, binds in the middle of the ORF, forward direction

CHAPTER 4: **Expanding the use of CRI-SPA by construction of a set of mating-competent diploid strains of *Saccharomyces cerevisiae***

Abstract

Currently and historically, yeast metabolic engineering has been concerned with a rather small number of *Saccharomyces cerevisiae* strains, most of which are of genetic backgrounds specifically adapted for easy handling in the laboratory. Strain-to-strain variation is thus often neglected in yeast cell factory design and the great genetic diversity within the species is largely unexploited. In this regard, CRI-SPA would be a great tool enabling the high-throughput introduction of a biosynthetic pathway to a large set of genetically diverse strains in order to assess their aptitude for production of various compounds. However, as most industrial and naturally occurring strains of *S. cerevisiae* are non-haploid and prototrophic, they are non-compatible with this method. Here, we employ a CRISPR-Cas9 based editing strategy to construct a small library of diploid *S. cerevisiae* strains of diverse origin that are able to mate and are compatible with the CRI-SPA method. A proof-of-principle test study further verified the expanded functionality of CRI-SPA for transfer of a gene deletion to the constructed, diploid strains. The presented approach thus represents a general strategy for constructing large libraries of genetically diverse strains compatible with the CRI-SPA screening procedure.

Introduction

Yeast cell factory research is a field which has typically paid little attention to the inherent genetic diversity found between different strains of *Saccharomyces cerevisiae*, as it has been dominated by the use of only a handful different laboratory strains. Amongst these, strains derived from the S288C and CEN.PK strain series are undoubtedly the most popular, a survey of research literature showing usage of the strains in more than 50% of metabolic engineering studies over the past 15 years (1,2). However, only between these two strain backgrounds, significant genetic differences can be observed and whole genome sequencing studies have for example identified more than 13 000 single nucleotide polymorphisms (SNPs) (3) and 3000 insertion/deletions (indels) between S288C and CEN.PK (4).

Comparative studies have also investigated the effect of using either of the two strain backgrounds on physiological behaviour and production performance of yeast cell factories. For instance, a study in which the two strains were engineered in an identical manner to produce vanillin glucoside, showed a 10-fold higher yield of the compound in S288C (5). Similarly, in the work leading up to the successful production of semi-synthetic artemisinin in yeast, clear differences were observed in fermentation performance of an S288C and CEN.PK strain engineered to produce the compound (6,7). Another study also confirmed marked differences in production of shikimic acid between S288C-derived strains and the two other commonly used strains YSG50 and INVSc1 (8). Furthermore, physiological differences between various strain backgrounds have been confirmed for a number of traits including plasmid maintenance (9), aptness of

hosting different heterologous pathways (10), the phenotypic effect of single gene deletions (11), and tolerance to antifungals (12).

As the mentioned studies clearly illustrate, the choice of strain background is of high importance when implementing a cell factory design. However, the heavy dependency on lab strains in metabolic engineering research is itself additionally problematic, because not only are these strains quite different from one another, they have also been adapted to life inside a laboratory and their genetic set-up and behaviour is remote from strains used for industrial scale fermentations. In order to simplify genetic engineering and to enable strain crossing, lab strains are typically made to be haploids of a stable mating type due to mutation of the *HO* gene. They are also frequently rendered auxotrophic for one or several essential nutrients to facilitate the use of auxotrophic selection markers. These modifications have a significant effect on cell metabolism and even when the mutated gene is complemented by the corresponding marker gene, the phenotype cannot be assumed to be identical to that of the prototrophic wild-type (13). Hence, although these strains are useful for laboratory experiments, observations and conclusions made from these experiments cannot easily be extrapolated to strains used in biotech and fermentation industry. Industrial strains are diploid, polyploid or aneuploid and show increased genetic diversity to strains used in the laboratory (14). They also exhibit greater robustness, as they can be cultivated under the harsh conditions of large-scale fermentation processes. Still, the genetic diversity of *S. cerevisiae* as a species is far from captured in the range of the widely used industrial strains.

S. cerevisiae is a ubiquitous organism, found not just in environments linked to human fermentation activities, but also in diverse ecological niches such as trees, soils and insects (15). These natural strains have a significantly higher degree of genetic diversity than their domesticated relatives (16–19). To get better sense of the scale of this diversity, one can observe the result of Wang *et al.* (20) as an illustrative example. In this study, strains of *S. cerevisiae* were isolated from a rainforest on Hainan Island in the south coast of China. When sequencing the genome of the isolates within this confined geographical region, the study could show a genetic diversity that was equivalent to that of the entire human population (20). With this great genetic diversity of *S. cerevisiae* in mind, it is fair to assume that the heavy dependency on just a few strains in metabolic engineering research is truly a limitation for the advancement of the field. At the same time, this presents exciting prospects for improving cell factory performance.

Numerous culture collections are established world-wide harbouring thousands of strains of *S. cerevisiae* (and other yeasts) isolated from diverse environments and industrial settings in different geographical locations (21,22). By making use of this largely unexploited resource, it is possible that improved production can be obtained by screening strains of various origin to find the most suitable host for the intended fermentation product. Additionally, by screening diverse populations, advantageous genotypes and phenotypes amongst the strain library can be studied, potentially highlighting new engineering opportunities for continued cell factory improvement. In this regard, CRI-SPA could be a useful tool for transferring a biochemical production pathway in high-throughput (HT) to a genetically diverse strain library, which can then be screened to assess the production performance of the different strains (Figure 1).

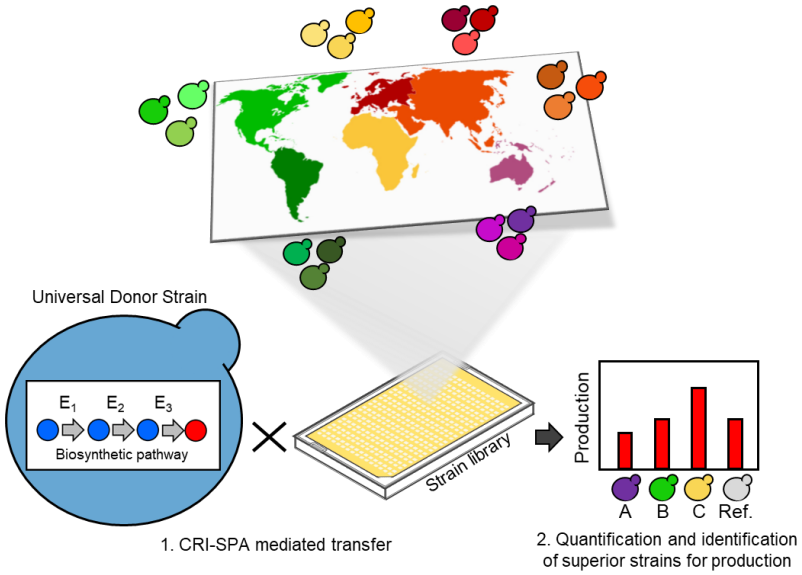


Figure 12. CRI-SPA-mediated transfer of a biochemical production pathway into a genetically diverse strain library. By making use of the great genetic diversity between strains of *S. cerevisiae*, superior host for cell factory-based production can be identified.

CRI-SPA is independent of meiosis and can be used with recipient strains that are non-isogenic to the Universal Donor Strain (UDS) without causing scrambling of the two genomes - an essential feature if one wishes to screen different genetic backgrounds. The method is, however, adapted for use with recipient strains that are haploid and *ura3*$^{-}$, as it requires mating to the UDS and a subsequent counter-selection step on 5-FOA for eliminating any remaining *URA3*-tagged chromosomes of the donor. This currently prevents the method from being used with industrially relevant strains or strains isolated from nature, which are typically non-haploid and prototrophic. As these are highly interesting from the perspective of exploiting genetic diversity for cell factory optimisation, it is an attractive pursuit to make them compatible with the CRI-SPA method to enable efficient and automated screening of their genetic backgrounds.

The mating type and mating behaviour in yeast is governed by the two mating loci *MAT***a** and *MATα*, consisting of distinct regions denoted X, Y, Z1 and Z2. The difference between the *MAT***a** and *MATα* allele lies within the Y region, which is denoted Y**a** and Yα, respectively. A *MAT***a** cell is able to mate to a *MATα* cell, whereas a diploid *MAT***a**/*MATα* is non-mating (23). A diploid cell can, however, be made to behave as either *MAT***a** or *MATα* by simply replacing one of the mating loci with the other to generate a diploid *MAT***a**/*MAT***a** or *MATα*/*MATα* cell, respectively. In 2018, Xie *et al.*, presented a system using CRISPR-Cas9, showing that diploid *S. cerevisiae* could efficiently be made mating-competent by designing gRNAs targeting either the *MAT***a** or *MATα* locus (24). As the loci differ in the 700 bp sequence of the Y**a** and Yα domain, they can be individually targeted and the generated double-strand break (DSB) at the locus can be repaired by using the homologous chromosome as template. In this process, the targeted *MAT* locus is replaced by the opposite mating type cassette, generating a diploid that is homozygous for the non-targeted mating locus and which is able to mate.

In the work presented here, we used the method described by Xie *et al.* to restore mating competency in a small set of industrial and natural isolates of *S. cerevisiae*. To ensure a stable mating type and compatibility with CRI-SPA, strains were further made *hoΔ ura3Δ*. Our experiments showed that the engineered strains were indeed able to mate and that they could be used in a proof-of-principle test case together with CRI-SPA for transfer of a genetic deletion. The work here thus demonstrates a general strategy for how large libraries of genetically diverse strains that are compatible with the CRI-SPA method can be generated.

Results

Generating mating-competent, CRI-SPA compatible diploid strains

The diploid strains that were chosen for this preliminary study are presented in Table 1 and are both of industrial and environmental origins. Strains are described and thoroughly characterized by (25) and were confirmed to have a diploid genome.

Table 5. Diploid strain backgrounds used in this study. Information is obtained from (25).

Strain name	Type	Isolation source	Geographic origin	Collection note
CA1	Industrial	Ethanol production	Brazil	Brazilian cachaca strain obtained from Rosane Schwan, Federal University of Lavras, Brazil
CBS7960	Industrial	Ethanol production	Brazil	Isolated from in a factory in Sao Paulo, Brazil and obtained from Dr. Justin Fay, Washington University School of Medicine. Produces ethanol from cane-sugar syrup.
CLIB382	Industrial	Beer brewing	Ireland	n.a
DBVPG1373	Environmental	Soil	The Netherlands	n.a
DBVPG6765	Environmental	Lychee fruit	Indonesia	n.a

In order to make the diploid strains compatible with the CRI-SPA method, several genetic modifications were required. First, strains needed to be made mating competent by replacement of one of their mating loci. For this purpose, the CRISPR-Cas9 based method developed by Xie *et al.* (2018) was used, targeting the *MATα* allele of the diploid strains in order to generate a *MAT**a**/MAT**a*** diploid able to mate to a *MATα* cell. The mating-type switching procedure is outlined in Figure 2A in which the targeted DSB in *MATα* is repaired using the *MATa* cassette on the homologous chromosome as repair template. In addition to the two active *MAT* loci, the yeast genome also harbours two silenced copies of the loci denoted *HMR**a*** and *HMLα* containing *MAT**a*** and *MATα* sequences, respectively. Due to the presence of heterochromatin structures mediated by *SIR* gene products at these loci (26), the two silenced mating cassette are not cleaved by endonucleases but may, although not illustrated in Figure 2A, serve as templates for homologous recombination in the repair of the DSB. However, the ultimate outcome would be the same: repair from *HMR**a*** would

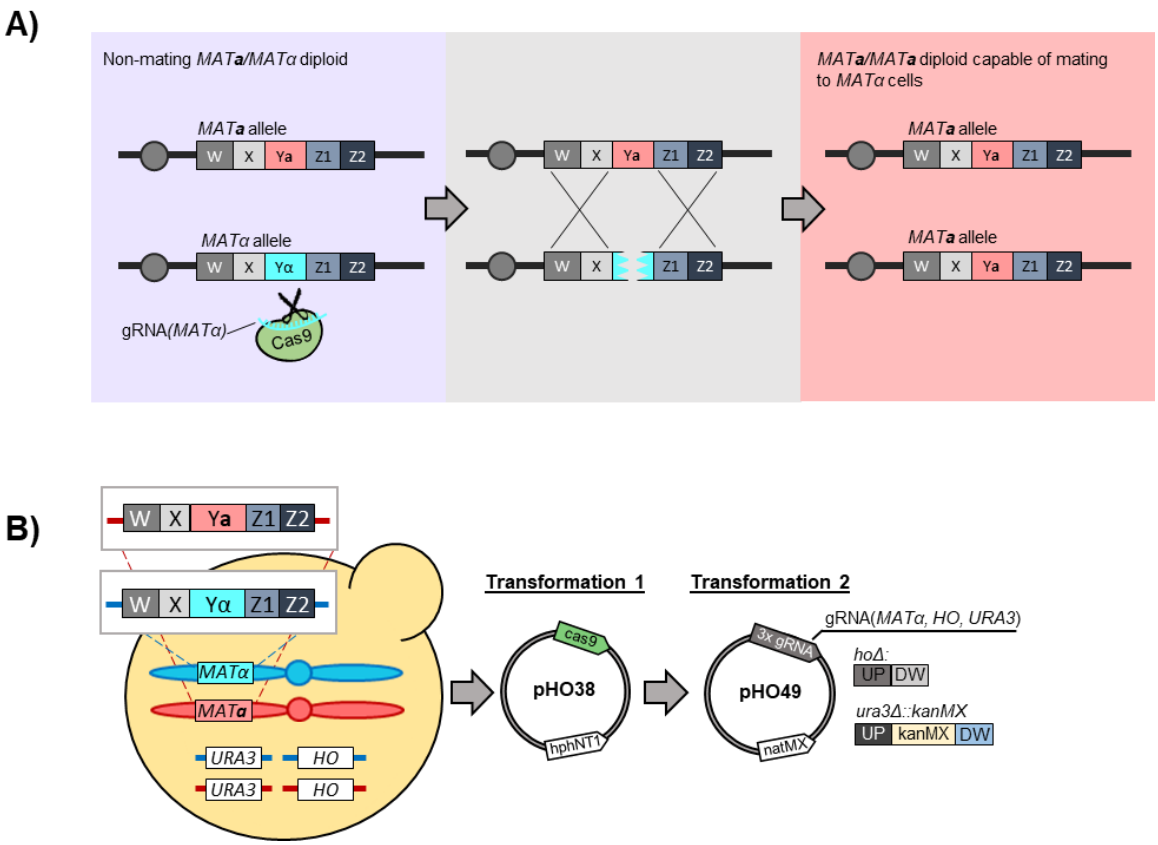


Figure 13. Employed strategy for generating CRI-SPA compatible diploid strains. A) CRISPR-Cas9 based strategy for restoration of mating competency in a diploid strain. By selective targeting of the *MATα* allele, homologous recombination using the homologous chromosome as repair template leads to loss of heterozygosity at the locus and yields a *MAT**a***/*MAT**a*** diploid capable of mating to *MATα* cells. Potential repair from *HMLα* or *HMRa* is not shown, see text for details. B) The complete procedure for generating mating-competent diploid strains of a stable mating type and which are lacking the *URA3* gene.. In a first transformation, the diploid strain is made to harbour the pHO38 Cas9 plasmid. In a second transformation, the strain is then co-transformed with a triple gRNA plasmid directing Cas9 cleavage of the Yα sequence of the *MATα* locus, the *HO* gene and the *URA3* gene, and linear repair substrates for the deletion of the two latter loci.

yield an identical result as illustrated in Figure 2A, and repair from *HMLα* would restore the *MATα* locus, causing Cas9 to cut the chromosome repeatedly until either the *MAT**a*** locus of the homologous chromosome or the *HMR**a*** locus is used to repair the break, thereby abolishing Cas9 cleavage.

In contrast to the study of Xie *et al.*, the diploid strains in this study had an intact *HO* gene. Therefore, in order to prevent mating type switching in *MAT**a***/*MAT**a*** strains, the *HO* gene additionally needed to be deleted. Furthermore, compatibility with the CRI-SPA method required deletion of *URA3* to enable the genome of the Universal Donor Strain to be counter-selected using 5-FOA. As there is one *MATα* allele, two alleles of the *HO* gene and two alleles of the *URA3* allele in a diploid genome, the overall strategy for generating CRI-SPA compatible strains required the simultaneous targeting of

five loci. Multiplex editing with CRISPR-Cas9 may be challenging, especially in industrial strains. We therefore chose to disrupt *URA3* by insertion of the *kanMX* cassette, increasing not only the efficiency of editing, but also making it possible to select for the genetic background of these modified strains and to use a similar plating protocol as that when using CRI-SPA with the yeast deletion library (Chapter 2 and 3 of this thesis).

The complete procedure for restoring mating competency and ensuring compatibility with CRI-SPA is outlined in Figure 2B. Strains were first transformed with the Cas9 plasmid pHO38 and were then, in a second round of transformation, co-transformed with the multiplex gRNA plasmid pHO49 targeting *MATα*, *HO* and *URA3* and linear deletion substrates *ho*Δ and *ura3*Δ::*kanMX.* In each case, in order to achieve sufficient transformation efficiency with these strains, it was necessary to increase the period of heat shock to 1 hour and to recover cells in rich medium overnight before plating transformed cells onto antibiotic-supplemented media. After selection on antibiotic plates, deletion of the two *URA3* alleles in streak-purified transformants was confirmed by replica plating onto SC-ura and by colony PCR in which deletion of *HO* alleles was also verified (data not shown). Although colony PCR also indicated the presence of only *MATa* loci in the constructed strains, their exact mating type required further confirmation. For this reason, verified *ho ura3* clones were put through a mating type test (Figure 3). Clones from the five strains were spotted onto YPD plates and, on these plates, strains were then mated to a *MATα* and *MAT**a*** tester strain. Control strains of known mating types were also included. After mating, cells were replica plated onto different antibiotic plates selecting either for the background of non-mated cells (kan^R *or* hyg^R) or of mated cells (kan^R *and* hyg^R). In each case, diploid strains were able to mate with the *MATα* tester strain, but not the *MATa* strain, indicating that a stable *MATa* mating type had been achieved.

CRI-SPA-mediated transfer of a gene deletion into constructed diploid strains

To verify compatibility of the constructed strains with the CRI-SPA method, the five strains were included in a small proof-of-principle experiment. When using CRI-SPA with diploid recipient strains, two alleles need to be targeted by Cas9 and although a marker is accompanying the transfer, there is no possibility to select for clones in which both chromosomes have been edited. Successful transfer is thereby dependent on the efficiency of Cas9 to cut non-edited chromosomes in order to eliminate heterozygous cells. To enable assessment of this efficiency, it was chosen to employ a CRI-SPA-based transfer of the *ade2*Δ::*hphNT1* allele (as in Chapter 2 of this thesis). Targeting of *ADE2* enabled direct indication of whether the system was sufficiently efficient to target both of its alleles in the diploid genome, as heterozygous clones will be resistant to hygromycin but will not display the red phenotype of a homozygous *ade2Δ*::*hphNT1/ade2Δ*::*hphNT1* cell.

The mating competent, diploid strains were arrayed on a high-throughput screening (HTS) agar plate in which each strain was added in quadruplicate in a 2 by 2 grid. By pin replication, the strains were then mated to the *Sc*_HO19 UDS containing the *ade2Δ*::hphNT1 deletion and harbouring a gRNA plasmid for Cas9-targeting of the *ADE2* locus (Figure 4A). Strains were taken through the CRI-SPA replica pining procedure outlined in Figure 4B, using growth media supplemented with elevated concentrations of adenine (40 mg/L) in order to lessen growth defects of *ade2* cells. In

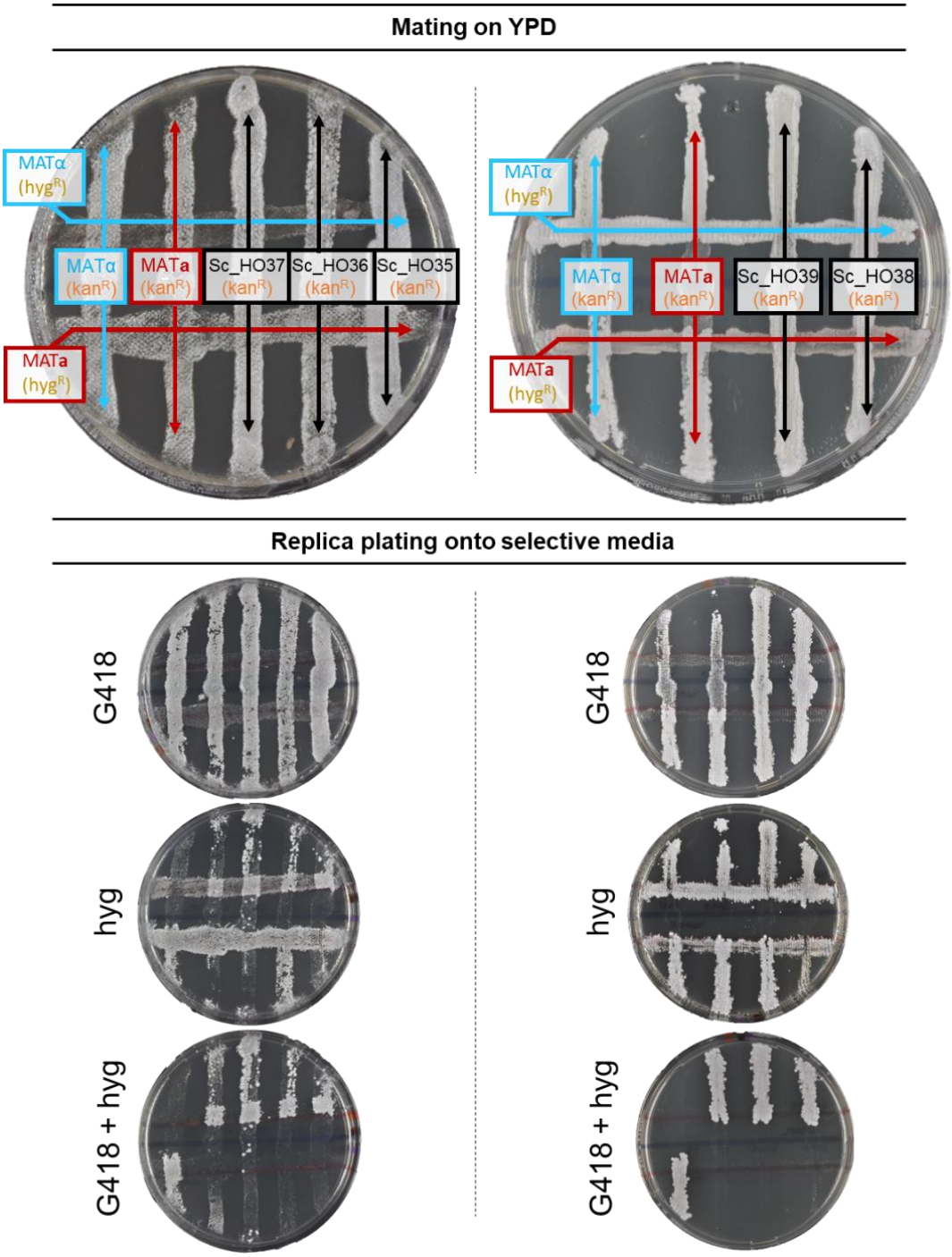


Figure 14. Verification of mating types in modified, diploid strains by mating to tester strains of known mating types and replica plating onto selective media. *MATα* hygR tester strain = *Sc*_HO19; *MAT**a*** hygR tester strain = BY4741 [pHO38]; *MATα* kanR control strain = C-VG, *MAT**a*** kanR control strain = *orfΔ::kanMX*, *Sc*_HO35 = CLIB383; *Sc*_HO36 = CA1; *Sc*_HO37 = CBS7960; *Sc*_HO38 = DBVPG1373; *Sc*_HO39 = DBVPG6765

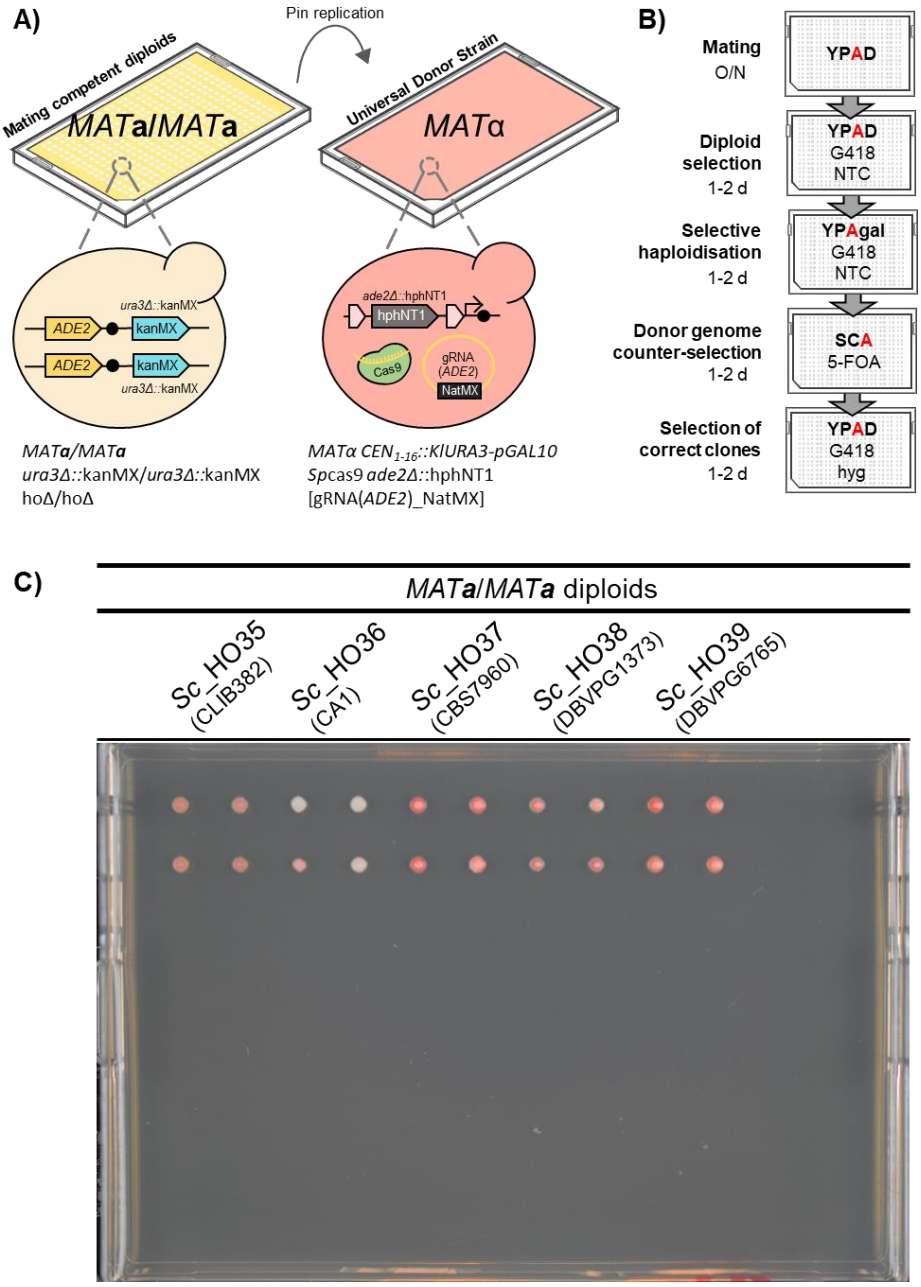


Figure 15. CRI-SPA- based transfer of a genetic deletion to a set of diploid strains rendered mating-competent and compatible with the CRI-SPA method.. (A) Experimental set-up. Diploid, recipient strains were arrayed on a rectangular agar plate and were then transferred by replica pinning onto a lawn of the Universal Donor Strain (UDS) *Sc*_HO19 in order to facilitate mating. The *Sc*_HO19 UDS harbours the *ade2Δ::hphNT1* allele and further carries a gRNA plasmid for Cas9 targeting of the wildtype *ADE2* locus.. (B) Employed pin replication protocol. Elevated concentrations of adenine (40 mg/L) was supplied to the media. (C) Final CRI-SPA generated strains after transfer onto synthetic solid medium with restricted concentrations of adenine (4 mg/L). Each strain is present in quadruplicate in a 2 by 2 grid).

contrast to previous CRI-SPA based transfers, mating was extended to overnight. After the final selection step, strains were transferred onto synthetic medium containing only 4 mg/L adenine in order to enhance the phenotype of *ade2* cells and the plate was photographed (Figure 4C). As can be observed in Figure 4C, for four out of the five strains, all four biological replicate displayed the expected red phenotype of a homozygous *ade2Δ* mutant and thus indicated efficient CRI-SPA mediated transfer of the gene deletion. For the fifth strain (*Sc*_HO36), Cas9 editing seemed to have been less efficient, but one of its replicates did display the red phenotype of a homozygous *ade2Δ* mutant and a second one appeared vaguely pink.

Discussion

The choice of strain background can have a significant effect on the performance of a cell factory, yet the great genetic diversity within the *S. cerevisiae* species has seen limited exploitation in yeast cell factory research. To tap into this resource, an automated HT procedure for transforming and assessing the production capacity for different heterologous compounds amongst libraries of diverse strains background would be highly desirable. World-wide culture collections harbour vast numbers of *S. cerevisiae* isolates from different geographical locations, environmental sources and human fermentation activities. These strains are unable to mate and may also have poor sporulation efficiencies, preventing the generation of haploid derivatives. We therefore found it worthwhile to investigate the possibility of making these non-haploid, non-mating strains compatible with the CRI-SPA method as it would enable facile, HT introduction of genetic material and subsequent screening of these genetically diverse strain backgrounds.

By using a CRISPR-Cas9 multiplex editing strategy, we simultaneously deleted the *MATα*, *HO* and *URA3* in a set of diploid strains of diverse origin. The generated strains were able to mate and, as discussed further below, were compatible with the CRI-SPA method. Feral and industrial isolates of *S. cerevisiae* pose a greater challenge in terms of genetic engineering. Not only are they more difficult to transform, but their ploidy also makes gene targeting more challenging. Our CRISPR-Cas9 targeting strategy entailed the simultaneous targeting of five genomic loci and did require some screening of transformants in order to find correct clones but was not unmanageable. Constructing a mating-competent and CRI-SPA compatible library of strain backgrounds would be a one-time task and the generated strains could then be included in any study aiming to improve production of a desired compound, simply by mating to different Universal Donor Strains.

When applying CRI-SPA in the transfer of a genetic feature into a diploid genome, correct editing of two loci is required. Using CRI-SPA with polyploid and aneuploid recipient strains would entail additional challenges. We tested the compatibility of our modified diploid strains with the CRI-SPA method by transfer of the *ade2*Δ deletion, which meant that correct deletion of both alleles could be determined by observing the phenotype of the resulting strains. We observed successful editing of both chromosomes in all four replicates in four of the five tested strains. The fifth strain yielded one replicate displaying the expected phenotype of an *ade2* clone. Although Cas9 editing was not 100% efficient, we found these initial results promising, demonstrating the feasibility of using the strains together with CRI-SPA. It is possible that

testing different protospacer sequences of the gRNA may be required for increased efficiencies, as small nucleotide variations in the genome of the recipient strains can have significant effects on the ability of the gRNA to direct Cas9 to its target site.

The aim of this study was to ultimately enable strains to be screened for their production capacity of different heterologous compounds and to use CRI-SPA in the transfer of a biochemical pathway (as outlined in Chapter 3 of this thesis). When transferring a biochemical pathway, it is not possible to determine whether genes have been integrated in only one or in both of the targeted chromosomal loci simply by observing the phenotype. It could be an interesting future prospect to construct recipient strains not only compatible with the CRI-SPA requirements, but which also make it possible to determine whether all targeted chromosomes have integrated the transferred DNA. With inspiration from the blue/white X-gal system in bacteria, it has been shown that a similar marker system is functional in yeast. The system makes use of the uidA gene, which enables yeast to break down X-gluc and to form a visibly blue compound (27,28). As recipient strains are modified in order to restore mating competence and ensure compatibility with the CRI-SPA method, it could be an attractive possibility to also introduce this marker in for example the XII-5 cloning site (29) (used in Chapter 3 of this thesis) of all chromosomes. In a CRI-SPA-based transfer, recipient strains like these would remain blue if harbouring non-modified loci when grown on X-gluc supplemented medium.

In previous studies using CRI-SPA, duplication of the recipient genome due to endoreduplication was observed – a phenomenon believed to be triggered by haploinsufficiency upon loss of chromosomes from a diploid genome. In this study, however, recipient strains are initially diploid and upon mating to the UDS form triploids before being brought back to a diploid state through selective loss of UDS chromosomes. The ploidy of the resulting CRI-SPA generated strains in this study was not investigated, although it would be highly interesting to see if similar endoreduplication processes are present in a 3n triploid converting to a 2n diploid.

In this study, the *URA3* alleles in the recipient strains were deleted in order to ensure compatibility with the CRI-SPA procedure. Auxotrophies can have significant impact on cell metabolism and is thus not ideal if one wishes to investigate the effect of genetic backgrounds on production capacities or other traits. A future possibility is to repress the expression of *URA3* rather than deleting the gene entirely. Transcriptional repression systems like the Tetracycline Off System (tetOff) (30) could be used to control the expression of *URA3* so that when grown on 5-FOA supplemented with tetracycline, the *URA3* gene is switched off, allowing growth of recipient strains but inhibiting any cells harbouring UDS chromosomes.

The work carried out in this study restored the mating competence of a set of diploid strains and ensured their compatibility with the CRI-SPA method in terms of *ura3* auxotrophy. However, in regards to the selective plating procedure of CI-SPA, one may also need to ensure that strains intended for use with the method are sensitive to 5-FOA and antibiotics used in the process. Although this was not a problem for the strains investigated in this study, variations in these traits are possible when diverse strains are being used. Additional aspects may also need to be taken into

consideration when working with genetically diverse strain backgrounds. Attention should be paid to the Convention of Biological Diversity and the Nagoya Protocol (https://www.cbd.int/). These international agreements are legally binding for signing countries and do for example hold rules for the commercial exploitation of microbes and for the equitable sharing of economic profits with countries from which the microbes originated. Furthermore, in addition to legal aspects, one should also bear in mind that opportunistic pathogenic *S.cerevisiae* strains exist and although mainly infecting immunocompromised individuals (31), their use requires awareness and proper safety precautions.

This work was a preliminary study including only a small number of strains. However, the results presented here demonstrate the feasibility of the chosen strategy for constructing CRI-SPA compatible strains and also verify their usage with the method. The strategy may thus easily be extended to other strains and can be used in the generation of larger strain libraries that when used together with CRI-SPA enables the genetic diversity of *S. cerevisiae* to be exploited in yeast cell factory development.

Methods

Strains and cultivation media

Strains were preserved in YPD glycerol stocks (20% (v/v)) at - 80°C. Throughout the study, strains were grown in YP (10 g/L yeast extract, 20 g/L peptone) using either glucose (20 g/l), galactose (20 g/l) or raffinose (20 g/L) as carbon source or in synthetic complete (SC) medium (prepared as described by (32) but with 60 mg/L L-leucine). For growth on solid medium 20 g/L agar was added. For selection, plates were supplemented with G418 (200 mg/L), nourseothricin (NTC) (100 mg/L) and hygromycin (hyg) (200 mg/L) in different combinations. When grown in liquid, only half of the concentration of the respective antibiotic was used. For selection on SC, the appropriate amino acid was dropped out. To enhance the phenotype of *ade2* cells, SC medium with only 4 mg/L adenine was prepared. To prevent *ade2* cells from reverting, growth media were supplemented with 40 mg/L adenine. For *URA3* counter selection, SC plates were supplemented with 5-fluoroorotic acid (5-FOA) (1 g/L) and uracil (30 mg/L). For cloning purposes, *Escherichia coli* DH5α was grown in LB medium (33) supplemented with 100 mg/L ampicillin and in the case for growth on solid medium, 20 g/L agar was added.

Diploid strains used in this study are presented in Table 1 along with information about their origin. All other strains used and constructed in this study are listed in Table 2.

Design of gRNA sequences

The gRNA sequence for targeting *MATα* was taken from Xie *et al.* 2018 (24). All other target sequences in this study were generated using the Benchling CRISPR tool (https://benchling.com). Off-target effects were minimized by using the S288C reference genome and the algorithm from Doench *et al.* 2016 (36).

Table 2. Other strains used and constructed in this study.

Strain name	Background	Relevant genotype	Source
*Sc*_HO19	W8164-2B	*MAT*α CEN$^{1\text{-}16}$::pGal1-*KlURA3* *ade2*Δ::p*Ag*TEF-hygR-tCYC1 X-3::p*TEF1-Sp*Cas9-t*CYC1*-loxP-*KlLEU2 KlURA3(*dw *ADE2)*	Chapter 2 of this thesis
BY4741	S288C	*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0	(34)
C-VG	CEN.PK	*MAT*α *adh6*Δ::kanMX	(5)
orfΔ::*kanMX*	BY4741	*MATa his3*Δ1 *leu2*Δ0 *met15*Δ0 *ura3*Δ0 *orf*Δ::*kanMX*	(35)
*Sc*_HO35	CA1	M*ATa/MATa ho*Δ *ura3*ΔkanMX	This study
*Sc*_HO36	CBS7960	M*ATa/MATa ho*Δ *ura3*ΔkanMX	This study
*Sc*_HO37	CLIB382	M*ATa/MATa ho*Δ *ura3*ΔkanMX	This study
*Sc*_HO38	DBVPG1373	M*ATa/MATa ho*Δ *ura3*ΔkanMX	This study
*Sc*_HO39	DBVPG6765	M*ATa/MATa ho*Δ *ura3*ΔkanMX	This study

General cloning procedures

Primers used in this study are presented in Supplementary Table 1 and were all ordered from Integrated DNA Technologies (Leuven, Belgium). All cloning was done using uracil-specific excision reagent (USER™) enzyme from New England Biolabs and was carried out as previously described (37,38). DNA polymerases were purchased from Thermo Scientific and used according to the supplier's instructions. Standard PCR amplifications were run using Phusion Hot Start II High-Fidelity DNA Polymerase. Amplification of DNA for USER cloning was carried out using Phusion U Hot Start DNA Polymerase. In the case of colony PCR, Taq DNA Polymerase was used. Prior to cloning, amplified DNA was treated with DpnI (Thermo Scientific) and was column or gel purified using the illustra GFX PCR DNA and Gel Band Purification Kit (GE Lifesciences). All plasmids were verified by sequencing at Eurofins MWG Operon (Germany) using primers specified in Supplementary Table 1.

Plasmid construction

All plasmid used and constructed in this study can be found in Table 3. The multiplex gRNA plasmid pHO49 targeting the *MAT*α, *HO* and *URA3* locus was constructed by first constructing plasmids targeting single loci. For this, the pCfB3050 plasmid was linearized by PCR using uracil-containing primers HOP60 and HOP61, excluding the target sequence of its gRNA cassette. Double-stranded DNA cassettes containing new 20 bp target sequences for the three different loci where then generated by annealing the following single-stranded oligo pairs: HOP209/HOP210 (*MAT*α), HOP382/HOP383 (*HO*) and HOP384/HOP385 (*URA3*). The annealing was done by mixing equal volumes of each oligo (100 µM), boiling for five minutes and then allowing to cool to room temperature. In each case, the double-stranded DNA cassettes contained single-stranded 5' and 3' overhangs and were individually USER cloned into the linearized pCfB3050 backbone, which upon USER treatment obtained matching 3'/5' overhangs to those of the cassettes. Plasmids were named pHO26 (gRNA(*MAT*α)), pHO27 (gRNA(*HO*)) and pHO40 (gRNA(*URA3*)). The double gRNA plasmid pHO37 was then constructed by amplifying the *HO* targeting gRNA cassette of pHO27 with primers HOP220

Table 3. Plasmids used and constructed in this study.

Plasmid name	Genetic element	Source
pCfB3050	p*SNR52*-gRNA(XII-5)-t*SUP4*, NatMX	(40)
pHO24	p*SNR52*-gRNA(*ADE2*)-t*SUP4*, NatMX	Chapter 2 of this thesis
pHO26	p*SNR52*-gRNA(*MATα*)-t*SUP4*, NatMX	This study
pHO27	p*SNR52*-gRNA(*HO*)-t*SUP4*, NatMX	This study
pHO37	p*SNR52*-gRNA(MATα, HO)-t*SUP4*, NatMX	This study
pHO38	pTEF1-cas9(*Hs*)-tCYC1, hphNT1	Chapter 3 of this thesis
pHO40	p*SNR52*-gRNA(*URA3*)-t*SUP4*, NatMX	This study
pHO49	p*SNR52*-gRNA(*MATα, HO, URA3*)-t*SUP4,* NatMX	This study

and HOP223 and cloning it into pHO26 linearized by primers HOP219 and HOP224. Finally, triple gRNA plasmid pHO49 was generated by amplifying the gRNA cassette of pHO40 with primers HOP220 and HOP221 and cloning into pHO37 linearized by HOP288 and HOP289.

Strain construction

Strains used and constructed in this study are listed and described in Table 2 above. All yeast transformations were carried out using the LiAc / SS Carrier DNA / PEG method (39) but with an extended time for heat shock of 1 hour. Additionally, cells were recovered overnight in YPD after heat shock before being plated onto antibiotic-supplemented plates. For all transformations, negative controls were included, where no DNA was added to the transformation mix. In each case, ca 200 ng plasmid DNA and between 700-1000 ng linear DNA was used. All linear DNA generated by PCR was gel purified prior to transformation. All strains was verified by running colony PCR to ensure correct integration at the intended chromosomal locus.

CRI-SPA high-throughput pin replication procedure

Automatic pin replication was carried out using the high-throughput pinning robot ROTOR HDA from Singer Instruments (United Kingdom), along with replica pinning pads (RePads) and rectangular petri dishes (PlusPlates) from the same company.

The generated CRI-SPA compatible diploid strains were arrayed on a rectangular agar plate, where each strain was present in quadruplicate in a 2 by 2 grid. This was done by first suspending the individual strains in 400 uL sterile water, which was then split into four wells of a 96 well microtiter plate and strains were thereafter transferred from the well plate onto the agar plate using the ROTOR HDA. For parallel mating of the Universal Donor Strain to the diploid library, a single colony of the *Sc*_HO19 UDS was inoculated in 5 mL YPD supplemented with NTC to maintain selective pressure for the gRNA plasmid. The UDS was grown overnight and 0.7 mL of the culture was then spread evenly on top of a rectangular YPD agar plate using sterile glass beads in order to generate a lawn of UDS cells. Once the lawn had dried in, the arrayed strains of the library was pinned on top and mating was allowed to proceed for 24 hours. Strains were

then pin replicated onto a series of selective plates as outlined in Figure 4A, each time allowing at least 24 hours of growth on each plate.

References

1. Hong KK, Nielsen J. Metabolic engineering of *Saccharomyces cerevisiae*: A key cell factory platform for future biorefineries. Cell Mol Life Sci. 2012;69(16):2671–90.
2. Kim IK, Roldão A, Siewers V, Nielsen J. A systems-level approach for metabolic engineering of yeast cell factories. Vol. 12, FEMS Yeast Research. 2012. p. 228–48.
3. Otero J, Vongsangnak W, Asadollahi MA, Olivares-Hernandes R, Maury J, Farinelli L, *et al.* Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications. BMC Genomics . 2010;11(1):723.
4. Nijkamp JF, van den Broek M, Datema E, de Kok S, Bosman L, Luttik MA, *et al.* De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. Microb Cell Fact . 2012;11(1):36.
5. Strucko T, Magdenoska O, Mortensen UH. Benchmarking two commonly used *Saccharomyces cerevisiae* strains for heterologous vanillin-??-glucoside production. Metab Eng Commun . 2015;2:99–108
6. Paddon CJ, Keasling JD. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Microbiol . 2014;12(5):355–67.
7. Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R. Production of amorphadiene in yeast , and its conversion to dihydroartemisinic acid , precursor to the antimalarial agent artemisinin. 2012;109(3):111–8.
8. Suástegui M, Guo W, Feng X, Shao Z. Investigating strain dependency in the production of aromatic compounds in *Saccharomyces cerevisiae*. Biotechnol Bioeng. 2016;113(12):2676–85.
9. Karim AS, Curran KA, Alper HS. Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. 2013;13:107–16.
10. Du J, Yuan Y, Si T, Lian J, Zhao H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. 2012;40(18):1–10.
11. Galardini M, Busby BP, Vieitez C, Dunham AS, Typas A, Beltrao P. The impact of the genetic background on gene deletion phenotypes in *Saccharomyces cerevisiae*. bioRxiv . 2018;44:487439.
12. Roberts CA, Miller JH, Atkinson PH. The genetic architecture in *Saccharomyces cerevisiae* that contributes to variation in drug response to the antifungals benomyl and ketoconazole. FEMS Yeast Res. 2017;17(3):fox027.
13. Pronk JT. MINIREVIEWS Auxotrophic Yeast Strains in Fundamental and Applied Research. Society . 2002;68(5):2095–100.
14. Steensels J, Snoek T, Meersman E, Nicolino MP, Voordeckers K, Verstrepen KJ. Improving industrial yeast strains: Exploiting natural and artificial diversity. FEMS Microbiol Rev. 2014;38(5):947–95.
15. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, *et al.* Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature. 2018;556(7701):339–44.
16. Winzeler EA, Castillo-davis CI, Oshiro G, Liang D, Richards DR, Zhou Y, *et al.* Genetic Diversity in Yeast Assessed With Whole-Genome Oligonucleotide Arrays. Genetics. 2003;163(1):79–89.

17. Dunn B, Levine RP, Sherlock G. Microarray karyotyping of commercial wine yeast strains reveals shared, as well as unique, genomic signatures. BMC Genomics. 2005;21:1–21.

18. Dunn B, Sherlock G. Reconstruction of the genome origins and evolution of the hybrid lager yeast Saccharomyces pastorianus. 2008;(650):1610–23.

19. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James S a, *et al.* Population genomics of domestic and wild yeasts. Nature. 2009;458(7236):337–41.

20. Wang QM, Liu WQ, Liti G, Wang SA, Bai FY. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. Mol Ecol. 2012;21(22):5404–17.

21. Boundy-Mills KL, Glantschnig E, Roberts IN, Yurkov A, Casaregola S, Daniel HM, *et al.* Yeast culture collections in the twenty-first century: new opportunities and challenges. Yeast. 2016;33(7):243–60.

22. Boundy-Mills K. Yeast culture collections of the world: Meeting the needs of industrial researchers. J Ind Microbiol Biotechnol. 2012;39(5):673–80.

23. Haber JE. Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. Genetics. 2012;191(1):33–64.

24. Xie Z, Mitchell LA, Liu H, Li B, Liu D, Agmon N. Rapid and Efficient CRISPR Cas9-Based Mating-Type Switching of *Saccharomyces cerevisiae*. G3. 2018;8(1):173–83.

25. Kang K, Bergdahl B, MacHado D, Dato L, Han TL, Li J, *et al.* Linking genetic, metabolic, and phenotypic diversity among *Saccharomyces cerevisiae* strains using multi-omics associations. Gigascience. 2019;8(4):1–14.

26. Hickman MA, Froyd CA, Rusche LN. Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage. Eukaryot Cell. 2011;10(9):1183–92.

27. Jefferson RA, Kavanagh TA, Bevan MW. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. 1987;6(12):3901–7.

28. Pretorius IS. Tailoring wine yeast for the new millennium: Novel approaches to the ancient art of winemaking. Yeast. 2000;16(8):675–729.

29. Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH, *et al.* Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metab Eng . 2012;14(2):104–11.

30. Garí E, Piedrafita L, Aldea M, Herrero E. A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cervisiae. Yeast. 1997;13(9):837–48.

31. Pérez-Torrado R, Querol A. Opportunistic strains of *Saccharomyces cerevisiae*: A potential risk sold in food products. Front Microbiol. 2016;6:1522.

32. Sherman F, Fink G., Hicks J. Laboratory Course Manual for Methods in Yeast Genetics. Cold Spring Harb Lab. 1986;

33. BERTANI G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. J Bacteriol. 1951;62(3):293–300.

34. Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast. 1998;14(2):115–32.

35. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science. 1999;285(5429):901–6.

36. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol . 2016;34(2):184–91.

37. Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA. USER fusion: A rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res. 2007;35(7):0–5.

38. Nour-Eldin HH, Hansen BG, Nørholm MHH, Jensen JK, Halkier BA. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res. 2006;34(18).

39. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc / SS carrier DNA / PEG method. Nat Protoc. 2008;2(1):31–5.

40. Jessop-Fabre MM, Jakočiūnas T, Stovicek V, Dai Z, Jensen MK, Keasling JD, *et al.* EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into *Saccharomyces cerevisiae* via CRISPR-Cas9. Biotechnol J. 2016;11(8):1110–7.

Supplementary information to Chapter 4

Supplementary Table 5 Primers used in this study. USER cloning tails are indicated in bold; gRNA protospacer sequences are underlined; long non-binding tails are written in lowercase letters.

Primer name	Sequence (5' to 3')	Purpose
Primers for cloning		
HOP60	**GTTTTAGAGCU**AGAAATAGCAAG	Linearizing plasmid pCfB3050 for cloning of new gRNA target sequence
HOP61	**GATCATTTAU**CTTTCACTGC	Linearizing plasmid pCfB3050 for cloning of new gRNA target sequence
HOP209	CAAATCATACAGAAACACAG**GTTTTAGAGCT**	Cloning gRNA target sequence for *MATα*, (+) strand
HOP210	CTGTGTTTCTGTATGATTTG**GATCATTTAT**	Cloning gRNA target sequence for *MATα*, (-) strand
HOP382	TATGGAAGATACAAATTCAG**GTTTTAGAGCT**	Cloning gRNA target sequence for *HO*, (+) strand
HOP383	CTGAATTTGTATCTTCCATA**GATCATTTAT**	Cloning gRNA target sequence for *HO*, (-) strand
HOP384	GGGTCAACAGTATAGAACCG**GTTTTAGAGCT**	Cloning gRNA target sequence for *URA3*, (+) strand
HOP385	CGGTTCTATACTGTTGACCC**GATCATTTAT**	Cloning gRNA target sequence for *URA3*, (-) strand
HOP219	**AGTGCAGGU**GCGATTAACTAATTACATGACTCGAAGAC	Cloning additional gRNA cassettes into single/double gRNA plasmid
HOP220	**ACCTGCACU**GCTGGAGCTTCTTTGAAAAGATAATG	Cloning additional gRNA cassettes into single/double gRNA plasmid
HOP221	**AGTCTGTGCAU**GCGATTAACTAATTACATGACTCGAAGAC	Cloning additional gRNA cassettes into single/double gRNA plasmid
HOP223	**AGCTGTGAGU**GCGATTAACTAATTACATGACTCGAAGAC	Cloning additional gRNA cassettes into single/double gRNA plasmid
HOP224	**ACTCACAGCU**CGCTCACTGACTCGCTG	Cloning additional gRNA cassettes into single/double gRNA plasmid
HOP288	**ATGCACAGACU**CCACAGAATCAGGGGATAACG	Cloning additional gRNA cassettes into single/double gRNA plasmid
HOP289	**AGTGCAGGU**GTATTACCGCCTTTGAGTGAGC	Cloning additional gRNA cassettes into single/double gRNA plasmid
Primers for sequencing		
HOP17	CGAATCGGACGACGAATCG	Sequencing gRNA plasmids
HOP171	**ATCGCGTGU**GTCACAGCTTGTCTGTAAGC	Sequencing gRNA plasmids
HOP198	**ATCGCGTGU**CGTCCATTCGCCATTCAGG	Sequencing gRNA plasmids
HOP224	**ACTCACAGCU**CGCTCACTGACTCGCTG	Sequencing gRNA plasmids
EDR441	AACCGTATTACCGCCTTTGAGTGAGCT	Sequencing gRNA plasmids
Primers for strain verification		
PJ0065	AGTCACATCAAGATCGTTTATGG	Verifying deletion of *MATα*
PJ0066	ACGGAATATGGGACTACTTCG	Verifying deletion of *MATα*
HOP257	CTTGAGATGGCGTATTTCTACTCCAG	Verifying deletion of *HO*
HOP305	CCTATGGTTTACGAAATGATCCACG	Verifying deletion of *HO*
PJ0059	TGCTGCAGATTTCTCCATCTCAC	Verifying deletion of *HO*
HOP276	CAGAAGGAAGAACGAAGGAAGG	Verifying insertion of kanMX at the *URA3* locus
HOP277	CGTCAGCCAGTTTAGTCTGACC	Verifying insertion of kanMX at the *URA3* locus
HOP278	GTCATTATAGAAATCATTACGACCGAGATTCC	Verifying insertion of kanMX at the *URA3* locus

CHAPTER 5: Discussion, perspectives and concluding remarks

The development of CRI-SPA was motivated by the need for a facile, high-throughput (HT) method for genetic modification of large yeast strain libraries, as these upon introduction of genetic designs and/or mutations can serve as important screening platforms in fundamental and applied research. Here, the ability of the method to fulfil its intended purpose is discussed. CRI-SPA is evaluated for its use in the different applications presented within this work. Previously made discussion points about its strength and limitations are recapitulated and expanded alongside new points for discussion, putting the method in context with other available strategies for genome-wide studies and metabolic engineering strategies. Future improvements and applications of CRI-SPA are addressed and finally, concluding remarks about CRI-SPA and the role of technology in the development of yeast cell factories are presented.

Putting CRI-SPA to the test

The mating-based, CRISPR-Cas9 assisted CRI-SPA method is highly amenable for use with the yeast deletion collection. By introduction of a second genetic deletion to the mutant strains of the library, the method allows studies of genetic interactions that may ultimately reveal new gene functions and cellular network structures (**Chapter 2**). We believe the method is also directly extendable to other mutant libraries including the GST-tagged ORF overexpression library (1), and the DAmP collection for studies of essential genes (2). In comparison to other mating-based methods for manipulation of the deletion collection, CRI-SPA benefits from its non-meiotic procedure, avoiding the complex and time-consuming process of sporulation which may otherwise decrease the throughput of the method. Furthermore, although not tested here, incorporation of CRISPR-Cas9 means that CRI-SPA could easily be used to transfer multiple mutations in tandem in order to study higher-order genetic interactions. In contrast to alternative methods such as SGA, CRI-SPA is independent of the co-segregation of modified chromosomes during meiosis, which enables truly high-level multiplexing as the method is not hindered by decreasing probabilities of chromosome co-segregation.

Our understanding of cellular metabolism is far from complete and living cells harbour complex regulatory systems and genetic interaction networks. This is a significant impedance to targeted cell factory optimisation. Screening of genome-wide mutation libraries to find strains with improved production of a target molecule can therefore be a viable alternative to rational metabolic engineering strategies. In addition to genetic interaction studies, CRI-SPA can also be used to introduce a biosynthetic pathway to the yeast deletion library to study the effect of each individual gene deletion on the ability to produce the compound in question. In this way, engineering targets that may be unintuitive from a rational metabolic engineering perspective can be identified and previously unknown functions of genes discovered. This was indeed the case when using CRI-SPA to transfer genes for the synthesis of betaxanthin to the mutant strains of the deletion collection (**Chapter 3**). The mutation leading to the highest production of betaxanthin in the employed screen was the deletion of *CSE2*, which positive impact appeared to be ascribed to its wide transcriptional regulatory functions rather than direct connection to any of the metabolic reactions of the aromatic amino acid pathway. The hypothesis does,

however, require further studies to gain mechanistic insight on the effects of the gene deletion on the phenotype. The screen also revealed a likely function of the *YOR1* gene product in exporting betaxanthin pigment to the extracellular environment, and the observation was supported by recent findings of another study on betaxanthin production in yeast (3). Furthermore, analysis of gene ontology (GO) terms amongst the top producers revealed that few mutations were linked to metabolic processes and instead showed an enrichment of genes connected to translational control. We did not verify the effect on phenotype of these mutations, but we note that these genes would have been difficult to identify as targets via rational metabolic engineering approaches and their effect on phenotype could not have been predicted by metabolic modelling.

Alternative methods for employing the deletion library in metabolic engineering studies exist. For example, traditional yeast transformation protocols may be used in HT set-ups but are cumbersome and when screening the *Saccharomyces* Genome Database for papers conducting genome-wide analyses in yeast, we could only find a few examples of when this approach was undertaken (4–6). Alternatively, the deletion library may also be pooled and transformed with DNA encoding the biosynthetic pathway of interest and top performing mutants selected and identified by barcode sequences (3). Recently, Cas9-gRNA based methods have also been presented which can be used to generate more or less systematic, genome-wide modifications to a background strain. The use of these methods involves transformation of the background strain with a pool of gRNA plasmids that target open reading frames (ORFs) across the genome and which enable their disruption (7) or their transcriptional reprogramming (when using catalytically dead Cas9 and CRISPR interference/activation) (8). The transformed cells are then screened for a phenotype of interest and the gRNA contained within top-scoring clones is identified. These CRISPR-Cas9 based strategies represent yet additional tools for conducting genome-wide studies in cell factory research and they offer simplicity and speed as the handling of large strain libraries are avoided. These and the methods for transforming the pooled deletion library do, however, have limitations in terms of screenable phenotypes. As resulting transformants need to be analysed in a pooled format, the studied phenotype need to allow either colorimetric or fluorescent evaluation for example by fluorescent activated cell sorting (FACS), or give a competitive growth advantage that allows enrichment of cells carrying beneficial mutations. In contrast, CRI-SPA is adapted for use with structured libraries and following transfer of a genetic feature into the library, the phenotype of individual mutant strains can be assessed. This means that, although time-consuming, analysis of compounds that do not produce a visible output or provide competitive advantage can be analysed by other methods, such as liquid chromatography. Analysis of production in individual strains directly from the agar plate may even be possible by the use of exciting, new HT analytical methods employing laser-enabled mass spectrometry (9)

In addition to the use with the genome-wide mutations libraries addressed above, CRI-SPA can also be used in conjunction with libraries of strains of different genetic backgrounds. Here, the functionality of CRI-SPA is unmatched by other mating-based methods for HT genetic modification, as these are dependent on meiotic processes for haploidisation, ultimately leading to shuffling of the two parental genomes. CRI-SPA is, however, meiosis-free and can be used with recipient strains that are non-isogenic to the Universal Donor Strain without causing genome shuffling. This

means that CRI-SPA can be used to transfer a biosynthetic pathway (or other genetic features) to a library of genetically diverse *S. cerevisiae* strains, which can then be screened for their production capacities. Although strains used in industry or strains isolated from the environment are not naturally able to mate, we have shown a feasible strategy for making diploid strains compatible with the method (**Chapter 4**). The future construction and use of large strain background libraries with CRI-SPA represents an exciting possibility for improving yeast cell factory performance by exploiting the great genetic diversity of the species.

Future perspectives

CRI-SPA is not without limitations and during the work presented in this dissertation, some key areas for improvement have been identified. Firstly, CRI-SPA is highly dependent on the ability of Cas9 to make correct edits to the genome of the recipient and to eliminate cells which fail to integrate the feature to be transferred. This is especially important as duplication of the recipient genome was observed in CRI-SPA generated strains (**Chapter 2**), which means that two recipient chromosomes need to be targeted in order to avoid generation of heterozygous cells. Furthermore, efficient Cas9 targeting is crucial for multiplex editing. Fortunately, the field of CRISPR-Cas9 is rapidly evolving and new CRISPR nucleases with novel characteristics and higher degrees of accuracy are frequently being discovered. However, numerous studies point to the abundance of gRNA as the key determinant of Cas9 editing efficiency rather than the nuclease itself. In the work presented here, the original set-up of DiCarlo *et al.* (10) was implemented, using the RNA polymerase III (pol III) p*SNR52* promoter to drive the expression of the gRNA. Other expression systems were found to improve Cas9 editing by employing Pol II promoters and by flanking the gRNA with self-cleaving RNA sequences such as ribozymes (11–13), tRNA sequences (14,15) or recognition sequences for the exogenous endoribonucleases Csy4 (16). These cleavage sequences release one or multiple functional gRNAs after expression, but also serve to protect transcripts from degradation in the cell (12). It is likely that the efficiency (in terms of correct and multiplex editing) of CRI-SPA would be benefitted by a switch to one of these alternative gRNA expression systems.

A second limitation is that CRI-SPA is dependent on the use of markers in its selective plating procedure. To ensure complete loss of donor chromosomes from mated and haploidised cells, donor chromosomes are marked with *URA3,* which enables counter-selection on 5-FOA but also adds the requirement of recipient strains being *ura*$^{-}$. In **Chapter 4**, an efficient system for deleting *URA3* in diploid strains was described and a future approach was also proposed in which expression of *URA3* is controlled by an inducible promoter in constructed strains. This would avoid the metabolism-altering effects of auxotrophy but allow compatibility with the CRI-SPA selective procedure. An alternative and possibly more convenient solution, would be to instead change the counter-selectable marker of the Universal Donor Strain. The dominant *amdSYM* marker is counter-selectable by growth on fluoroacetamide-supplemented medium (17) and replacement of the *URA3* sequences in the donor genome with this marker would allow CRI-SPA to be used with prototrophic recipient strains without any prior modifications. Additionally in terms of markers, although efficient Cas9 targeting allows marker-free editing, the use of markers is likely needed when using CRI-SPA for multiplex editing and, when possible, the use of markers also allows a general robustness to HT methodologies. It is therefore possible that

the availability of markers can limit CRI-SPA based screens and, as discussed in **Chapter 3**, it would be highly interesting to investigate the implementation of split-enzyme-marker systems in CRI-SPA (18).

Finally, screening is a main bottleneck in any HT method generating a large number of DNA constructs, protein variants or, in the case of CRI-SPA, strains with distinct genetic profiles. Recently, development of a growing number of genetically encoded biosensor for a range of compounds makes it possible for the producing microbes themselves to report their production capacities by producing a measurable output, commonly in the form a fluorescent signal. As addressed in **Chapter 3**, there are now several biosensors that can be used to directly report production of a target compound in yeast (19) and which could be incorporated in CRI-SPA, transferring not only the biosynthetic pathway for the compound in question, but also the biosensor to report on its production. The continued development and expansion of biosensor to detect additional compounds in yeast should be a useful contribution to future CRI-SPA based screens.

Limitations aside, this study has served to establish the functionality and usefulness of CRI-SPA and looking into the future, additional applications are also possible. Some of these have already been discussed in previous chapters (including higher-order genetic interaction studies and iterative cycles of metabolic engineering screens) and a few additional ones are now also addressed here.

In order for cell factories to be a truly sustainable alterative to oil-based production, they need to operate using non-food biomass that can be generated without negative environmental impact such as for example deforestation for clearing of agricultural land. In addition to improving and expanding the portfolio of molecules produced by microbial systems, it is therefore also key to develop cell factories that can utilise carbon sources derived from non-food based feedstocks, such as for example lignocellulosic waste residues from forestry and agriculture, and waste products from industry and households. It would be highly relevant to use CRI-SPA not only for transfer of biosynthetic genes but also for heterologous substrate utilisation pathways for non-conventional sugars, e.g. xylose (20) that is found in lignocellulosic materials, or mannitol (21) that is abundant in seaweed biomass. The library could then be used to screen for strain backgrounds or mutations leading to improved growth on these substrates.

In addition to being used with mutant and genetically diverse strain libraries, construction of other CRI-SPA compatible libraries could be relevant. These strain libraries could for example contain strains expressing different molecular chaperones (or combinations thereof). This would allow studies on the effect of protein folding capacities on the production of a target compound, perhaps most interestingly those which are synthesised by heterologous enzymes that can be suspected to be inefficiently processed in their non-native host.

Finally, there are six relatives of *S. cerevisiae* within the *Saccharomyces sensu stricto* genus: *S. paradoxus, S. mikatae, S. kudriazevii, S. arboricola, S. eubayanus and S. uvarum.* Interspecies crossing between these organisms is possible but resulting hybrid diploids are typically sterile, unable to produce viable spores (22). However, since CRI-SPA is independent of meiosis, it is possible that the method in the future may be used with libraries harbouring strains of these

different *Saccharoymces* species. This is an exciting prospect as these organisms display vast genetic diversity, their protein sequence divergence being on par with that between birds and humans (23).

Concluding remarks

This work was initiated with the aim of providing CRI-SPA as a novel tool in the continued search and improvement of yeast cell factories. This dissertation will now finish with a general reflection on the role of this and other technologies in the development of metabolic engineering and its possibilities of providing a biobased economy with microbial cell factories that offer viable and sustainable production of chemicals.

Since the advent of recombinant DNA technology and molecular cloning in the 1970s, the field of biological engineering has seen great technological development. The cost for synthesising DNA is continuously decreasing and standardised methods for DNA assembly has sped up cloning procedures. In parallel, synthetic biology is developing engineering-inspired principles for supplying 'off-the-shelf', modular DNA parts with well-characterised, specific biological functions that offer scalability and orthogonality to genetic designs. Moreover, DNA sequencing is faster and cheaper, not only allowing labs to more efficiently verify their constructs and strains, but also making the genetic information available for an ever increasing number of organisms, in turn enabling new genes and enzymes with interesting properties to be discovered. Altogether with the elucidation of genomes, expanded genetic toolboxes for non-conventional yeasts and other microbes makes it increasingly possible to harness these as microbial platforms for production of various compounds.

With regards to genetic toolboxes, the relatively recent impact of CRISPR-Cas9 on the speed and precision of genome editing has been immense. The technology is likely only at the start of its development and new applications are continuously found allowing for example transcriptional reprogramming (24), nucleotide base editing (25), epigenetic edits of DNA methylation (26), and cellular imaging (27).

In addition to gene editing being faster thanks to technologies such as CRISPR, task in the laboratory are increasingly streamlined by implementation of automated procedures enabling HT cloning and strain construction (28). Furthermore, the development of systems biology has enabled mathematical modelling of cellular metabolism and *in silico* predictions of phenotypes. These models are getting increasingly sophisticated and the genome-scale metabolic model for *S.cerevisiae* does for example incorporate thermodynamic properties of reactions and abundance data of biomolecules such as enzymes and lipids in order to constrict the model outputs to better agree with biology and to improve predictive capacities. (29–31). In parallel, biology is increasingly driven by data, where the collection of large sets of genomic, transcriptomic, proteomic and metabolomics data is aiding systems biology to better model living systems. Additionally, the data is also used as training sets in the development of machine learning technologies that can make predictions and decisions in situations where detailed understanding of underlying biological mechanisms are lacking (32).

CRI-SPA is a humble contribution to the continued technological development in the field. The method offers novel and unique possibilities for high-throughput introduction of genetic modifications to large, structured yeast strain libraries,

exploiting the genetic information contained within these libraries and facilitating their use in both fundamental and applied science. The combination of facile mating-based transformation with the multiplex and versatile functional capacities of CRISPR-Cas9 should allow CRI-SPA to be used in the study of diverse research questions.

Yet, despite the fast development and continued addition of new tools and technologies in genetic and metabolic engineering, only a small number of yeast cell factories and other microbial production systems are today operating at a commercial scale. The challenges in the development of bio-based production processes are not only technological and the success of a microbial cell factory is not only dependent on the ability of metabolic engineers to optimise its metrics. Despite its harmful effects on the environment, human health and the climate, oil continues to be cheap and the cost of emitting CO_2 low. The fossil fuels industry is additionally receiving significant tax subsidies from national governments, which within the EU alone reached around €110 billion between the years 2014 and 2016 (33). Considering the narrow time-window within which drastic decreases of greenhouse gas emissions need to be achieved in order to prevent the most severe effects of climate change, political change and regulations are needed to limit the use of fossil resources and to promote sustainable alternatives. Such political incentives together with the continued technological advance and development in metabolic engineering should make yeast cell factories and other microbial production systems key components of a future sustainable bioeconomy.

References

1. Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, *et al.* Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell. 2006;21(3):319–30.
2. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat Methods. 2008;5(8):711–8.
3. Savitskaya J, Protzko RJ, Li FZ, Arkin AP, Dueber JE. Iterative screening methodology enables isolation of strains with improved properties for a FACS-based screen and increased L-DOPA production. Sci Rep. 2019;9(1):1–10.
4. Kitagawa T, Kohda K, Tokuhiro K, Hoshida H, Akada R, Takahashi H, *et al.* Identification of genes that enhance cellulase protein production in yeast. J Biotechnol . 2011;151(2):194–203.
5. Hirasawa T, Takekuni M, Yoshikawa K, Ookubo A. Genome-wide identification of the targets for genetic manipulation to improve l- lactate production by *Saccharomyces cerevisiae* by using a single-gene deletion strain collection. J Biotechnol . 2013;168(2):185–93. A
6. Özaydin B, Burd H, Lee TS, Keasling JD. Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. Metab Eng. 2013;15(1):174–83.
7. Bao Z, HamediRad M, Xue P, Xiao H, Tasan I, Chao R, *et al.* Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. Nat Biotechnol. 2018;36(6):505–8.
8. Ferreira R, Skrekas C, Hedin A, Sánchez BJ, Siewers V, Nielsen J, *et al.* Model-Assisted Fine-Tuning of

Central Carbon Metabolism in Yeast through dCas9-Based Regulation. ACS Synth Biol. 2019;8(11):2457–63.

9. Gowers GOF, Cameron SJS, Perdones-Montero A, Bell D, Chee SM, Kern M, *et al.* Off-Colony Screening of Biosynthetic Libraries by Rapid Laser-Enabled Mass Spectrometry. ACS Synth Biol. 2019;8(11):2566–75.

10. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. 2013;41(7):4336–43.

11. Gao Y, Zhao Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. J Integr Plant Biol. 2014;56(4):343–9.

12. Ryan OW, Skerker JM, Maurer MJ, Li X, Tsai JC, Poddar S, *et al.* Selection of chromosomal DNA libraries using a multiplex CRISPR system. Elife. 2014;3(2014):e03703.

13. de Vries ARG, de Groot PA, van den Broek M, Daran JMG. CRISPR-Cas9 mediated gene deletions in lager yeast Saccharomyces pastorianus. Microb Cell Fact . 2017;16(1):1–18.

14. Xie K, Minkenberg B, Yang Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc Natl Acad Sci . 2015;112(11):3570–5.

15. Nødvig CS, Hoof JB, Kogle ME, Jarczynska ZD, Lehmbeck J, Klitgaard DK, *et al.* Efficient Oligo nucleotide mediated CRISPR-Cas9 Gene Editing in Aspergilli. Fungal Genet Biol . 2018;(January).

16. Ferreira R, Skrekas C, Nielsen J, David F. Multiplexed CRISPR/Cas9 Genome Editing and Gene Regulation using Csy4 in *Saccharomyces cerevisiae*. ACS Synth Biol . 2017;7(1):10–5.

17. Solis-Escalante D, Kuijpers NGA, Bongaerts N, Bolat I, Bosman L, Pronk JT, *et al.* amdSYM, A new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. FEMS Yeast Res. 2013;13(1):126–39.

18. Jillette N, Du M, Zhu JJ, Cardoz P, Cheng AW. Split selectable markers. Nat Commun . 2019;10(1).

19. D'Ambrosio V, Jensen MK. Lighting up yeast cell factories by transcription factor-based biosensors. FEMS Yeast Res. 2017;17(7):1–12.

20. Li X, Chen Y, Nielsen J. Harnessing xylose pathways for biofuels production. Curr Opin Biotechnol. 2019;84(3):e02146-17.

21. Enquist-Newman M, Faust AME, Bravo DD, Santos CNS, Raisner RM, Hanel A, *et al.* Efficient ethanol production from brown macroalgae sugars by a synthetic yeast platform. Nature. 2014;505(7482):239–43.

22. Hittinger CT. Saccharomyces diversity and evolution: A budding model genus. Trends Genet . 2013;29(5):309–17.

23. Dujon B. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. Trends Genet. 2006;22(7):375–87.

24. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell . 2013;154(2):442–51.

25. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature . 2016;533(7603):420–4.

26. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, *et al.* Editing DNA Methylation in the Mammalian Genome. Cell . 2016;167(1):233-247.e17.

27. Ma H, Tu LC, Naseri A, Huisman M, Zhang S, Grunwald D, *et al.* Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. Nat Biotechnol. 2016;34(5):528–30.

28. Chao R, Mishra S, Si T, Zhao H. Engineering biological systems using automated biofoundries. Metab Eng .

2017;42(June):98–108.

29. Soh KC, Miskovic L, Hatzimanikatis V. From network models to network responses: Integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. FEMS Yeast Res. 2012;12(2):129–43.

30. Sánchez BJ, Li F, Kerkhoven EJ, Nielsen J. SLIMEr: Probing flexibility of lipid metabolism in yeast with an improved constraint-based modeling framework. BMC Syst Biol. 2019;13(1):1–9.

31. Sánchez BJ, Zhang C, Nilsson A, Lahtvee P, Kerkhoven EJ, Nielsen J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol Syst Biol. 2017;13(8):935.

32. Yu R, Nielsen J. Big data in yeast systems biology. FEMS Yeast Res. 2019;19(7):1–10.

33. Gençsü I, Mclynn M, Runkel M, Burg L Van Der, Worrall L, Whitley S, *et al.* Phase-out 2020 : Monitoring Europe's Fossil fuel Subsidies . 2017.